

## THESIS / THÈSE

### MASTER EN SCIENCES INFORMATIQUES À FINALITÉ SPÉCIALISÉE EN DATA SCIENCE

#### Visualisation des déplacements domicile-travail des agents de la ville de Charleroi

Maquoi, Jerome

*Award date:*  
2020

*Awarding institution:*  
Universite de Namur

[Link to publication](#)

#### General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

#### Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.



**UNIVERSITÉ  
DE NAMUR**

---

FACULTÉ  
D'INFORMATIQUE

**Visualisation des déplacements  
domicile-travail des agents de la ville de  
Charleroi**

Jérôme MAQUOI



## Résumé

En ce début de XXI<sup>e</sup> siècle, la mobilité urbaine est un enjeu crucial dans nos sociétés. En effet, les villes doivent apporter des réponses aux nombreuses problématiques qui y sont liées, et qui prennent toujours plus d'importance. La ville de Charleroi ne fait pas exception. Dans ce contexte, elle aimerait s'inscrire dans une mobilité intelligente, notamment vis-à-vis des déplacements domicile-travail de ses agents, c'est à dire les personnes travaillant pour la ville. L'objectif de ce mémoire est de comprendre comment des techniques de visualisation peuvent aider la ville de Charleroi à connaître les déplacements domicile-travail de ses agents et à leur proposer un mode de transport adapté à leurs déplacements et le plus durable possible. Pour ce faire, ce mémoire commence par synthétiser la littérature des techniques de visualisation de trafic existantes. Il décrit ensuite le développement d'une proposition de solution répondant à la problématique puis discute des perspectives d'amélioration de cette solution. Finalement, ce mémoire espère fournir des outils à la ville de Charleroi afin de l'aider à développer des solutions informatiques qui répondent à des problématiques de mobilité auxquelles elle fait face.





## Remerciements

Je tiens tout d'abord à remercier mon promoteur, le Professeur Bruno Dumas, et mon co-promoteur, le Doctorant Antoine Clarinval, qui m'ont épaulé tout au long du stage et de la rédaction de ce mémoire. Ils étaient tous deux disponibles pour répondre à toutes mes questions. Leurs conseils et leurs remarques m'ont permis d'augmenter la qualité de mon mémoire.

Ma reconnaissance va aussi à ma Maître de Stage Pascaline Gilson. Malgré son emploi du temps chargé, elle m'a toujours prêté une oreille attentive et a su se rendre disponible lorsque j'avais besoin d'elle. Elle m'a également permis de gagner en autonomie en me faisant confiance tout au long du développement de la solution. Pour tout cela, je la remercie chaleureusement.

Merci aussi à Cathy Libois et Géraldine Ansotte qui, avec Pascaline Gilson, ont accepté de jouer le rôle du client. Leurs remarques m'ont été très utiles et m'ont permis de développer une solution la plus pertinente possible.

J'aimerais aussi remercier l'ensemble des membres du cabinet Mobilité de la ville de Charleroi pour leur accueil chaleureux. J'ai tout de suite été intégré dans l'équipe comme l'un des leurs et ils ont toujours veillé à ce que je me sente bien durant tout mon stage.

Enfin, je remercie mes parents et Clémence Hooreman pour leur soutien, spécialement durant le confinement, et les corrections apportées sur ma rédaction.



# Table des matières

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Etat de l’art</b>	<b>3</b>
2.1	Mobilité . . . . .	3
2.1.1	Définition de la mobilité . . . . .	3
2.1.2	Les grands thèmes de la mobilité à Charleroi . . . . .	4
2.2	Visualisation . . . . .	5
2.3	Analyse spatiale . . . . .	6
2.3.1	Définition . . . . .	6
2.3.2	Niveaux de mesures . . . . .	6
2.3.3	Symboles géographiques et niveaux de mesure . . . . .	7
2.3.4	Types de données et de cartes . . . . .	8
2.4	Visualisation du trafic . . . . .	8
2.4.1	Objectifs de la visualisation au niveau de la mobilité . . . . .	9
2.4.2	Techniques de visualisation de trafic . . . . .	10
2.4.3	Croisement des objectifs et des techniques de visualisation . . . . .	19
2.5	Zoom sur les <i>flow maps</i> . . . . .	23
2.5.1	Simplification spatiale . . . . .	23
2.5.2	<i>Edge bundling</i> . . . . .	24
2.5.3	Autres techniques de réduction de confusion . . . . .	26
2.6	Conclusion . . . . .	28
<b>3</b>	<b>Développement de la solution</b>	<b>29</b>
3.1	Méthodologie . . . . .	29
3.2	Données dont la Ville dispose . . . . .	30
3.2.1	Données du Système d’Information Géographique . . . . .	30
3.2.2	Données d’analyse de trafic de FLOWCheck . . . . .	31
3.2.3	Données de déplacement de Proximus . . . . .	32
3.2.4	Données de la police de Charleroi . . . . .	32
3.2.5	Données d’utilisation des transports en commun . . . . .	33
3.2.6	Données sur les agents de la Ville . . . . .	33
3.2.7	Données KML provenant de cartes Google Maps . . . . .	34
3.3	Inventaire des besoins et difficultés rencontrées . . . . .	35
3.3.1	Gestion des parkings . . . . .	35
3.3.2	Confrontation de FLOWCheck avec les données radars de la police . . . . .	35
3.3.3	Confrontation de FLOWCheck avec les choix proposés par les GPS . . . . .	36
3.3.4	Affinage des résultats de FLOWCheck . . . . .	36
3.3.5	Simplification de l’utilisation de FLOWCheck . . . . .	37
3.3.6	Création d’une carte interactive de mobilité . . . . .	37
3.3.7	Gestion des déplacements domicile-travail des agents . . . . .	38
3.4	Développement d’une solution concernant les déplacements domicile-travail des agents . . . . .	39
3.4.1	Analyse de la problématique . . . . .	39
3.4.2	Données en temps réel provenant d’API . . . . .	40
3.4.3	Architecture . . . . .	42
3.4.4	<i>Preprocessing</i> des données . . . . .	44
3.4.5	Développement des visualisations . . . . .	46

3.4.6	Coût de la solution proposée . . . . .	59
3.4.7	Évaluation de la solution proposée . . . . .	60
3.5	Discussion sur les contributions . . . . .	62
<b>4</b>	<b>Perspectives d'amélioration</b>	<b>64</b>
4.1	Introduction . . . . .	64
4.2	Visualisations à destination des expert la Ville . . . . .	64
4.2.1	Intégration des visualisations dans QGIS . . . . .	64
4.2.2	Ajout de nouvelles données . . . . .	64
4.2.3	Affinage des visualisations . . . . .	65
4.3	Visualisation à destination des agents de la Ville . . . . .	65
4.3.1	Ajout d'une interface d'administration . . . . .	65
4.3.2	Ajout d'un paramétrage de la solution par l'utilisateur . . . . .	66
4.3.3	Changement de la cible . . . . .	66
4.3.4	Affinage de la visualisation . . . . .	66
<b>5</b>	<b>Conclusion</b>	<b>68</b>

# Chapitre 1

## Introduction

La mobilité est un sujet fort discuté dans nos sociétés. En effet ce domaine se décline en plusieurs thématiques qui deviennent de plus en plus importantes en ce début de XXI<sup>e</sup> siècle : le nombre de voitures en circulation dans le monde, le nombre d'avions civils en vol, les routes commerciales internationales, les infrastructures liées à la mobilité comme les routes et les rails de chemin de fer ou encore les mouvements de population dûs à des migrations qui peuvent être écologiques, politiques, religieuses ou économiques. Ces migrations touchent particulièrement les villes où la mobilité a une grande influence sur le mode de vie des habitants. En effet, selon les Nations Unies (United Nations, 2014), plus de la moitié de la population mondiale vivra en ville en 2050. Ce phénomène a déjà un impact majeur sur la mobilité urbaine et cet impact sera de plus en plus fort dans les années à venir.

La ville de Charleroi n'est pas épargnée par cet exode rural. Dans ce contexte, elle aimerait s'inscrire dans une mobilité intelligente pour faire face à cet afflux de nouveaux habitants prévu pour les années à venir. Au cours des dernières décennies, la mobilité n'a pas été mise en avant dans la gestion de la ville. Par conséquent, elle possède encore peu de données de mobilité intéressantes et ces données sont très souvent statiques alors que la mobilité est un phénomène dynamique en perpétuelle évolution. Ces dernières années, la ville a changé son fusil d'épaule. Elle commence à se donner les moyens, via l'acquisition de jeux de données, de faire face aux nombreux besoins de mobilité qui sont dorés et déjà présents mais aussi d'anticiper les besoins futurs.

La mobilité à Charleroi englobe plusieurs thèmes comme la gestion du trafic, des événements organisés en ville, des budgets, de la voirie ou encore de la sécurité routière, pour ne citer qu'eux. La gestion du trafic est sans doute un des thèmes les plus importants en mobilité car elle impacte directement et en profondeur les habitudes de vie des citoyens ainsi que la mobilité en ville dans sa globalité. Au sein de cette gestion du trafic, la ville fait face à différentes problématiques qui nécessitent des réponses appropriées.

Une de ces problématiques concerne les déplacements domicile-travail des personnes travaillant pour la ville. Ces travailleurs peuvent être par exemple des enseignants, des secrétaires ou encore des personnes travaillant dans l'urbanisme ou l'état civil. La ville a l'obligation de répondre régulièrement à un questionnaire fédéral sur les déplacements de ses travailleurs. Pour ce faire, elle doit connaître les habitudes de déplacement de l'ensemble de ses travailleurs, ce qui n'est pas le cas jusqu'à présent. De plus, toujours dans le contexte des déplacements domicile-travail, la ville aimerait pousser ses travailleurs à utiliser des moyens de transport plus durables. Le besoin est donc double : connaître les habitudes de déplacement domicile-travail des personnes travaillant pour la ville et proposer à ces travailleurs un moyen de transport plus durable.

Afin d'y répondre, la ville de Charleroi doit recueillir un maximum de données. En effet, il n'est pas possible de réaliser une analyse pertinente si elle ne connaît rien de la situation. Cependant, posséder des données ne suffit pas. Il faut pouvoir les exploiter et leur donner du sens. Malheureusement, il est très compliqué pour des experts d'effectuer une analyse sur

des énormes masses de données. L'humain ne sait pas les appréhender car les données sont trop volumineuses.

Un moyen de résoudre ce problème est l'introduction de visualisations. La visualisation permet de représenter visuellement les données afin d'aider les experts à les appréhender et à les analyser. Avec ces analyses, les experts peuvent par après répondre au besoin défini. C'est en quelque sorte un pont entre les données et les besoins. L'objectif de ce mémoire est de construire ce pont et plus précisément de comprendre quelles techniques de visualisation peuvent aider à répondre au besoin défini ci-dessus. La question de recherche de ce mémoire est la suivante :

Quelles techniques de visualisation de l'information peuvent aider la ville de Charleroi à connaître les déplacements domicile-travail de ses agents et à leur proposer un mode de transport adapté à leurs déplacements et le plus durable possible ?

Afin de proposer des solutions répondant à ce problème, un stage s'est déroulé à l'hôtel de ville de Charleroi durant le deuxième quadrimestre de l'année 2019-2020, sous la supervision de Pascline Gilson, attachée Mobilité au cabinet Mobilité de la ville de Charleroi. Le stage a débouché sur le développement de deux outils de visualisation. Le premier aide la ville à visualiser les flux de déplacement de ses agents. Le second est une plateforme à destination de chaque agent de la ville permettant à ceux-ci de visualiser et comparer les différents moyens de transport pour effectuer les trajets entre leur domicile et leur lieu de travail. Ces deux outils ont été développés à partir de la littérature scientifique tout en suivant le *framework* **WHAT WHY HOW**. Ce *framework* consiste à d'abord analyser les données disponibles (**WHAT**), puis analyser un ou plusieurs besoins du client (**WHY**) et enfin proposer une visualisation qui répond à ce ou ces besoins en fonction des données disponibles (**HOW**).

Ce mémoire est organisé comme suit. Le chapitre 2 décrit un état de l'art de techniques de visualisation potentiellement intéressantes pour répondre à la question de recherche. Le chapitre 3 décrit le développement de la solution proposée. Comme le développement a suivi le *framework* **WHAT WHY HOW**, il a donc été divisé en trois parties. La première discute des données dont la ville de Charleroi dispose. La deuxième présente les besoins identifiés au cours du stage. L'un de ces besoins est à la base de la définition de la question de recherche de ce mémoire. Enfin la troisième partie discute de l'implémentation des visualisations répondant à la question de recherche. Le chapitre 4 décrit les perspectives d'amélioration de la solution proposée. Enfin, le chapitre 5 conclut ce mémoire.

# Chapitre 2

## Etat de l'art

La question de recherche s'insère dans le domaine de la mobilité. L'état de l'art commence donc par définir ce concept dans la section 2.1. Une fois le thème bien défini et discuté, la notion de visualisation est elle-même définie dans la section 2.2, suivie du principe d'analyse spatiale, qui va guider la recherche de techniques de visualisation, à la section 2.3. Ensuite l'état de l'art se centrera sur le trafic et les techniques de visualisation le concernant à la section 2.4. Enfin, les techniques utilisées pour développer une solution répondant à la question de recherche seront discutées à la section 2.5.

### 2.1 Mobilité

#### 2.1.1 Définition de la mobilité

(Larousse, sd) définit la mobilité comme suit : "Propriété, caractère de ce qui est susceptible de mouvement, de ce qui peut se mouvoir ou être mû, changer de place, de fonction" Cette définition est très générale mais donne un aperçu d'une notion importante : la notion de mouvement. Le Larousse va plus loin et ajoute comme définition, dans le domaine de la sociologie : "Changement de lieu de résidence (mobilité géographique), ou d'activité économique (mobilité professionnelle)".

La mobilité fait donc référence à un changement qui peut s'effectuer dans plusieurs domaines : le Larousse parle de mobilité géographique ou professionnelle mais ce ne sont pas les seuls types de mobilité. Plusieurs auteurs scientifiques donnent leur définition de la mobilité qui complète celle vue ci-dessus. Urry (2007) développe le concept en le divisant en cinq sous-domaines : la mobilité des objets, la mobilité physique, la mobilité imaginative, la mobilité virtuelle et la mobilité communicative. La mobilité des objets se rapporte au déplacement d'objets entre des personnes. La mobilité physique ou corporelle discute du mouvement de personnes. La mobilité imaginative quant à elle est l'idée de la mobilité telle que définie à travers différents médias comme par exemple la presse, la radio ou le cinéma. La mobilité virtuelle fait référence à la mobilité de ce qui se trouve sur Internet. Enfin la mobilité communicative discute des communications entre personnes liées à du mouvement.

Fortunati et Taipale (2017) se positionnent par rapport à Urry (2007) en amenant une nouvelle définition de la mobilité plus centrée sur l'humain. L'article propose une nouvelle manière de voir le concept à travers différents sous-domaines : la *macro-mobility*, la *micro-mobility*, la *media-mobility* et la *disembodied-mobility*. Tout d'abord, la *macro-mobility* fait référence aux mouvements de beaucoup de personnes sur de longues distances contrairement à la *micro-mobility* qui s'intéresse aux plus petits mouvements, à petite échelle. La *media-mobility* parle de la responsabilité des smartphones dans la nouvelle mobilité des médias traditionnels ainsi que la manière dont ces derniers se sont intégrés dans les moyens de transport. Enfin la *disembodied-mobility* traite des différentes modifications sociales de notre société.

Les définitions de Urry (2007) et de Fortunati et Taipale (2017) concernant la mobilité



ne sont pas exclusives entre elles. En effet, la *macro-mobility* et la *micro-mobility* ajoutent une notion d'échelle de mesures qui peuvent être utilisées avec les types de la définition de Urry (2007).

### 2.1.2 Les grands thèmes de la mobilité à Charleroi

Dans le cadre de la ville de Charleroi, appelée la Ville dans la suite de ce document, la gestion de la mobilité prend des formes diverses. Voici une liste non exhaustive des différents domaines compris dans la mobilité et sur lesquels la Ville réfléchit afin de résoudre des problèmes ou améliorer les services déjà existants. Cette liste est issue de la liste des dossiers mobilité de la ville de janvier 2020.

Un premier dossier de mobilité est la politique de la Ville en matière d'automobile, de vélo et de transport en commun. Ces trois moyens de transport sont discutés en long et en large afin d'améliorer la mobilité des riverains. Par exemple, au niveau cyclable, la Ville met en place des plans vélos, réfléchit aux moyens qui peuvent être mis en oeuvre pour augmenter le nombre de cyclistes tant au point de vue incitatif qu'au niveau des infrastructures qu'il faudrait développer, et met en place des sens uniques limités, sens uniques ouverts dans les deux sens pour les cyclistes. Au niveau automobile, la Ville travaille sur la création de parkings aux endroits clefs ou encore sur les taxes Viapass à destination des poids-lourds pour gérer leur passage en ville. Enfin, au niveau des transports en commun, la Ville réfléchit principalement à une nouvelle ligne de métro dans le sud de la Ville ainsi qu'au projet Bus à Haut Niveau de Service permettant à des bus de circuler sur des routes non accessibles en voiture.

Discuter de la voiture, du vélo et des transports en commun amène à la multi modalité, c'est-à-dire l'utilisation de plusieurs moyens de transport pour se rendre quelque part. Dans cette optique, la Ville réfléchit à la création de parkings de covoiturage ainsi qu'à conclure des partenariats avec la TEC pour offrir des abonnements aux citoyens afin d'augmenter l'attractivité des transports en commun.

Les thèmes discutés ci-dessus sont les plus importants. Cependant il est intéressant de noter que ce ne sont pas les seuls. Il y a aussi par exemple la cartographie qui joue un grand rôle en mobilité. Elle regroupe principalement les cartes présentant les informations de circulation dont la Ville dispose comme par exemple son infrastructure : zones bleues, zones 30, écoles, stationnement, ... Un autre exemple de thème est la gestion des lieux d'intérêt c'est-à-dire les lieux particuliers qui sont intéressants au niveau de la mobilité comme les places ou les écoles. Pour ces dernières, des plans de sécurisation sont développés par la Ville afin de mettre en place une infrastructure pour protéger un maximum les élèves des voitures ou des bus passant aux alentours des écoles.

La gestion de la mobilité au sein d'une grande ville comme Charleroi est une tâche ardue. En effet, de nombreux paramètres doivent être pris en compte pour définir une politique cohérente et qui fait sens. De plus, des problèmes perturbent et compliquent cette gestion.

Le premier est le manque de personnes travaillant dans ce domaine. L'histoire de la Ville a fait que la mobilité n'était pas une priorité durant les dernières décennies. La conséquence est donc le nombre restreint de personnes travaillant dans le domaine de la mobilité de la première ville de Wallonie en terme de population qui atteint plus de 200 000 habitants en 2020.

Le deuxième concerne l'évolution de la mobilité. Ces cinquante dernières années, la mobilité urbaine s'est transformée à plusieurs reprises sous l'impulsion de plusieurs facteurs comme des événements qui changent radicalement la manière de se déplacer des citoyens ou les politiques mises en place à Charleroi. L'affaire Dutroux en 1996 est un exemple d'événement qui a changé la manière de se déplacer des élèves du secondaire de la région de Charleroi. Avant cette triste affaire, la plupart des élèves allaient à l'école en transports en commun. Par après, les élèves continuant à prendre les transports en commun n'étaient plus que 20%. Cette habitude est restée jusqu'à aujourd'hui, ce qui explique cette grande diffé-

rence entre les habitudes de transport des élèves de Charleroi par rapport à d'autres villes comme par exemple Liège où environ 80% des élèves vont à l'école en transports en commun.

De manière générale, concernant les politiques mises en place durant ces cinquante dernières années, Jones (2014) explique qu'elles sont passées par trois étapes. La première étape est la focalisation sur la croissance de l'utilisation de la voiture. Une conséquence est la construction d'infrastructures routières. Cependant celles-ci ne parviendront à gérer ni l'afflux de voitures ni des problèmes de congestion et de pollution.

La deuxième étape part de ce constat d'inefficacité des infrastructures. Là où la première étape se centrait sur la voiture, cette étape-ci porte son attention sur les personnes. Comment faire pour aller le plus rapidement possible d'une origine à une destination ? Les politiques se concentrent d'avantage sur l'utilisation de moyens de transport alternatifs comme les transports en commun car ils sont plus efficaces en ville que les voitures personnelles. Cette attention sur les personnes va être de plus en plus forte et mener à une troisième étape.

Celle-ci se concentre sur la qualité de vie au sein des espaces urbains ainsi que la vision des villes comme des endroits regroupant l'activité économique. Dans cette optique, des politiques sont mises en oeuvre afin de réduire la pollution en ville, augmenter l'utilisation des transports en commun et diminuer celle de la voiture personnelle, promouvoir des activités de rues et améliorer la qualité de vie des citoyens.

La transformation de la mobilité au fil de ces trois étapes révèle quelque chose d'important pour la ville de Charleroi. L'évolution de la mobilité est très rapide et doit s'adapter en permanence à de nouveaux paradigmes. Cette adaptation implique que les personnes travaillant dans ce domaine doivent rapidement prendre des décisions très importantes. Afin de prendre les meilleures décisions, il est important d'avoir le plus d'informations possibles sur la situation de la Ville. C'est ici qu'intervient un problème de quantité de données. Soit il y en a trop et il est difficile de les appréhender et de sortir l'utile de l'accessoire, soit il n'y en a pas assez et il est difficile d'avoir une vision de ce qu'il se passe.

## 2.2 Visualisation

Munzner (2014) définit la visualisation comme l'expression visuelle d'ensembles de données ayant pour but d'aider des personnes à réaliser un travail. De manière similaire, Chen et al. (2015) et Azzam et al. (2013) la définissent comme une méthode permettant de modifier les données brutes et de les représenter de manière visuelle afin d'aider l'utilisateur à les comprendre et les analyser. En effet, lorsque de trop grandes quantités de données doivent être analysées, le cerveau humain ne sait pas comment les traiter car elles sont trop nombreuses. Par contre, celui-ci a plus de facilités à appréhender et analyser des données ou de l'information provenant de représentations visuelles (McDaniel et Einstein, 1986) (MIT, 1996). Dans ce contexte, la visualisation fait le pont entre les capacités de calcul de l'ordinateur et les capacités de réflexion de l'Homme et permet de définir quelle utilisation faire des données, comment leur donner du sens et comment les utiliser dans des processus de prise de décision. C'est pour toutes ces raisons que la visualisation est très importante dans la gestion des données.

Pour Chen et al. (2015), la visualisation de données se divise en trois principaux domaines. Tout d'abord, la visualisation scientifique permet de représenter spatialement des données relatives à de la physique ou de la chimie. A ce titre, Munzner (2014) propose une définition de la visualisation scientifique qui vient compléter celle de Chen et al. (2015). Pour lui, la visualisation scientifique est la représentation de données pour lesquelles une position spatiale est connue. Ensuite, la visualisation de l'information se concentre sur la représentation de données complexes, c'est-à-dire ayant un grand nombre de propriétés ou dimensions (Ware, 2020). Enfin, le troisième domaine est l'analyse visuelle. Elle est une conséquence des deux premiers (Wong et Thomas, 2004) mais ajoute de l'interaction et des procédés permettant une meilleure analyse des données présentes sur la visualisation (Chen et al., 2015). Elle inclut également des techniques venant d'autres secteurs comme l'analyse statistique ou les théories de la décision afin de réaliser des analyses les plus complètes possible.

Pour d'autres auteurs, l'analyse visuelle a une définition très proche de celle de la visualisation de données de Chen *et al.* (2015). En effet, Andrienko *et al.* (2017) définit l'analyse visuelle comme "la science du raisonnement analytique facilité par des interfaces visuelles interactives". Elle permet à l'Homme et l'ordinateur de collaborer sur des tâches, chacun d'eux apportant une expertise différente. L'ordinateur est très puissant pour effectuer des calculs et traiter des données. L'Homme ne sait pas réaliser autant de calculs. En revanche, il est beaucoup plus puissant que l'ordinateur lorsqu'il s'agit de réfléchir et analyser des données. Ainsi, la puissance de calcul de l'ordinateur est utilisée pour traiter les données. Ces traitements sont ensuite envoyés à l'Homme qui les analyse et prend des décisions en fonction des résultats de son analyse. Quelle que soit la définition retenue, les interfaces Homme-machine sont primordiales. Elles doivent aider les utilisateurs à explorer et analyser les données. Dans la suite de cet état de l'art, l'ensemble des articles traités se basent sur la définition de l'analyse visuelle formulée dans la littérature comme une extension de la visualisation de l'information et qui apporte une notion d'interaction entre la visualisation et l'utilisateur.

## 2.3 Analyse spatiale

### 2.3.1 Définition

Au niveau de la mobilité et dans le cadre d'une ville, la plupart des données utilisées sont basées sur des cartes géographiques. En effet, les analyses faites dans ce domaine ont besoin de données concernant la position d'une série d'éléments qui forment une infrastructure. Cette infrastructure permet l'existence de la mobilité urbaine. Voiries, abris de bus, gares, rails, itinéraires des transports en commun sont autant d'exemples d'éléments qui constituent cette infrastructure et qui nécessitent des données géographiques pour les identifier, les traiter et les analyser afin de prendre des décisions. Les analyses faites sur ces données doivent donc prendre en compte cette composante spatiale.

Ce type d'analyse est appelé "analyse spatiale" et regroupe plusieurs démarches utilisées dans plusieurs situations (O'sullivan et Unwin, 2014) :

1. Manipulation de données spatiales via notamment des systèmes d'information géographiques (SIG) qui permettent de rassembler, traiter et analyser des données de type géographique (Maliene *et al.*, 2011). Ces systèmes sont implantés en profondeur dans les processus de représentation et de manipulation de données spatiales à travers le monde.
2. Analyse de données spatiales permettant l'exploration et la description des données. Une telle analyse est par exemple nécessaire lorsque l'analyse a affaire à un nouveau *dataset*. Avant de réaliser quelque chose avec ce *dataset*, il doit analyser ce qu'il contient.
3. Analyse statistique spatiale qui, comme son nom l'indique, utilise des techniques statistiques pour examiner les données disponibles dans le but de définir un modèle statistique. Par exemple, des données sur des accidents de la route pourraient être analysées de cette manière pour définir un modèle statistique sur les types d'accidents.
4. Modélisation spatiale lorsque l'objectif est de prédire l'évolution de données comme par exemple le trafic routier dans une zone particulière pour les prochaines années.

Ces démarches s'entremêlent tant et si bien qu'il est difficile de les dissocier et de parler de l'une sans l'autre. En effet, il est possible de manipuler des données pour les analyser et leur donner du sens. Par après, des modèles statistiques pourraient être construits sur ces analyses afin de réaliser des modélisations spatiales permettant de prédire l'évolution de ces données.

### 2.3.2 Niveaux de mesures

Afin d'obtenir le plein potentiel de l'analyse spatiale, il est nécessaire de recueillir le plus d'informations possible des données disponibles. Pour ce faire, Stevens *et al.* (1946) a défini une classification des données en quatre niveaux ou échelles : *nominal*, *ordinal*, *interval* et *ratio*. Ces différents niveaux permettent de préciser la nature des données et sont définis de

manière incrémentale, c'est-à-dire que chaque niveau ajoute une fonctionnalité au niveau précédent. Dans l'ordre du plus simple au plus complexe, Stevens et al. (1946) définissent les niveaux de mesure comme suit :

1. *Nominal* : mesure la plus simple donnant un label à une donnée permettant de la classer. Les données *nominal* ne sont définies que par cette classification et n'ont donc pas de lien avec une quelconque valeur quantitative.

Par exemple, le sexe d'une personne peut être classé dans la catégorie "féminin" ou "masculin". Cette donnée n'est pas quantitative et est juste représentée par cette classification.

2. *Ordinal* : mesure permettant de classer les données comme la mesure *nominal* ainsi que de leur donner un ordre. Cet ordre implique que la donnée a une valeur quantitative. Par exemple, une ville a un niveau de population parmi "petite", "moyenne" ou "grande", la population étant la valeur quantitative de la donnée. C'est aussi une manière de classer les données mais à la différence de la mesure *nominal*, cette mesure-ci donne une classification sur base de relations du type "sont plus grandes que", "sont plus petites que" ou "sont égales à".

3. *Interval* : mesure utilisée pour des données ayant des valeurs numériques, dont l'ordre (*ordinal*) et le type (*nominal*) sont connus, pour montrer la relation entre les différentes valeurs numériques qui sont relatives les unes aux autres. Ce niveau de mesure affine la mesure *ordinal* en permettant de comparer les variables entre elles. En effet la valeur des écarts entre chaque variable est connue.

Par exemple, des données représentant la température en Celsius sont de la mesure *interval* car leur type, la température, et leur ordre, du plus petit au plus grand ou l'inverse, sont connus. De plus, il est possible de comparer les températures entre elles comme par exemple  $+10^{\circ}\text{C}$  et  $+15^{\circ}\text{C}$  qui ont une différence de  $5^{\circ}\text{C}$  l'une par rapport à l'autre.

Cependant cette mesure a des limites car il n'est pas possible de calculer de ratio entre les différentes variables. En effet celles-ci ne possèdent pas de zéro absolu sur lequel elles peuvent se baser. Par exemple  $0^{\circ}\text{C}$  ne signifie pas l'absence de température. Bien qu'il soit possible de dire que  $10^{\circ}\text{C} + 10^{\circ}\text{C} = 20^{\circ}\text{C}$  et que donc  $20^{\circ}\text{C}$  représente une température deux fois plus grande que  $10^{\circ}\text{C}$ , lorsque ces températures sont converties en degrés Fahrenheit, cette comparaison tombe. En effet  $10^{\circ}\text{C} = 50^{\circ}\text{F}$  et  $20^{\circ}\text{C} = 68^{\circ}\text{F}$ , or  $68^{\circ}\text{F}$  n'est pas deux fois plus grand que  $50^{\circ}\text{F}$ . Pour pouvoir effectuer des analyses de ratio, il faut introduire une dernière mesure : la mesure *ratio*.

4. *Ratio* : mesure fournissant l'information la plus détaillée sur une donnée. Elle est utilisée sur des données dont le type (*nominal*), l'ordre (*interval*) et des valeurs numériques (*interval*) sont connus et ajoute une notion de zéro absolu à la donnée. Ce zéro absolu permet d'effectuer des analyses poussées de la donnée en la comparant avec d'autres données via des ratios.

Par exemple, des données sur le poids de personnes sont de type *ratio* car elles ont un type, le poids en kg, un ordre et des valeurs numériques en plus d'avoir un zéro absolu représenté par le poids nul, zéro kg, qui signifie une absence de poids. Ce poids nul permet de comparer différents poids entre eux et de dire par exemple qu'une personne pèse 1.25 fois le poids d'une autre personne.

### 2.3.3 Symboles géographiques et niveaux de mesure

La définition de ces mesures permet de réaliser une comparaison avec les différents symboles utilisés sur des cartes : les points, les lignes et les aires. Un point est une position géographique. Il est donc caractérisé par des coordonnées et peut donner une information sur une caractéristique de la donnée présente à ces coordonnées. Une ligne est définie comme un ensemble de points qui, mis bout à bout, forment un trait continu. Une aire définit un ou plusieurs attributs pour un ensemble de points présents dans une zone géographique spécifique (Wenzhan et al., 2003).

La figure 2.1 montre une classification entre les symboles et les différentes mesures définies ci-dessus avec pour chaque croisement entre mesure et symbole un exemple de donnée

(Robinson et Randall, 1969). Cette figure est intéressante car elle joint les différents niveaux de mesure aux principaux symboles utilisés en cartographie, ce qui permet de voir des exemples concrets montrant ce qu'il est possible de réaliser en fonction du type de données disponibles.

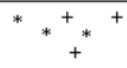
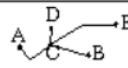


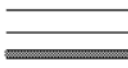

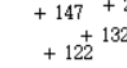



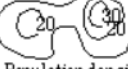
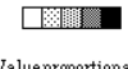
Content scaling level	Defining relations	FORM OF CARTOGRAPHIC SYMBOL		
		POINT	LINE	AREA
<b>Nominal</b>	Equivalence	 Wholesale and retail establishments	 Highway connectivity	 Land ownership
<b>Ordinal</b>	Equivalence Greater than	 Small Medium Large Population centers	 Roads by degree of improvement	 Crop yield
<b>Interval</b>	Equivalence Greater than Ratio of intervals	 Spot elevations	 Graticule	 Date of settlement
<b>Ratio</b>	Equivalence Greater than Ratio of intervals Ratio of scale values	 Area proportional to population	 Population density isopleths	 Value proportional to population density

FIGURE 2.1 – Classification par niveau de mesure (*scaling*) et symbole géographique (d'après (Robinson et Randall, 1969)).

### 2.3.4 Types de données et de cartes

Unwin (1981) complète ce que Robinson et Randall (1969) présentent dans la figure 2.1 par deux classifications, toujours par rapport aux niveaux de mesures et aux symboles cartographiques, avec pour ces derniers l'ajout du volume. La première classification concerne le type de données qui peuvent être représentées. Elle donne les mêmes informations que la figure 2.1 avec l'ajout d'informations concernant le volume. De plus, elle agrège les mesures *interval* et *ratio*. La deuxième classification donne le type de carte pouvant être utilisée dans chaque cas. Les deux classifications sont présentées dans la figure 2.2.

## 2.4 Visualisation du trafic

La question de recherche se base sur un sous-domaine de la mobilité : le trafic. En effet, les notions de déplacement domicile-travail ainsi que de moyens de transport impliquent une discussion préalable du trafic afin de répondre au mieux à la question de recherche. Chen et al. (2015) définissent le trafic comme "le flux ou le passage de véhicules motorisés, non motorisés, et de piétons sur la route, ou le mouvement de passagers". Cette notion de mouvement explique pourquoi le trafic est un sous-domaine de la mobilité. Cependant il est spécifique au mouvement de personnes ou de véhicules sur des voies prévues à cet effet, comme des routes, contrairement à la mobilité qui s'intéresse au mouvement en général.

Dans ce contexte, la section 2.4.1 explique les objectifs poursuivis par la visualisation de la mobilité en général. Ces objectifs s'insèrent aussi dans la visualisation du trafic. Les techniques de visualisation du trafic sont discutées à la section 2.4.2 et la comparaison entre les objectifs et les techniques est présentée à la section 2.4.3.

#### DATA TYPES

	Point	Line	Area	Volume
Nominal	City	Road	Name of unit	Precipitation or soil type
Ordinal	Large city	Major road	Rich county	Heavy precipitation Good soil
Interval	Total population	Traffic flow	Per capita income	Precip. in mm Cation exchange
Ratio				

#### MAP TYPES

	Point	Line	Area	Volume
Nominal	Dot map	Network map	Colored area map	Freely colored map
Ordinal	Symbol map	Ordered network map	Ordered colored map	Ordered chromatic map
Interval	Graduated symbol map	Flow map	Choropleth map	Contour map
Ratio				

FIGURE 2.2 – Types de données et types de cartes (d'après (Unwin, 1981)).

### 2.4.1 Objectifs de la visualisation au niveau de la mobilité

La visualisation de l'information peut apporter de nouveaux points de vues sur des données. Ces points de vue peuvent être de l'ordre de l'exploration ou de la prédiction de données. Cependant, pour qu'ils soient pertinents, la visualisation doit répondre correctement à des objectifs. Les objectifs présentés dans cette section sont valables pour la gestion du trafic mais aussi pour la mobilité en général. Il est important de les garder en tête lors du développement d'une solution qui se base sur une visualisation de l'information car ils définissent la plus-value que la visualisation peut apporter aux problématiques de mobilité. Chen et al. (2015) et Gondim et al. (2014) listent les objectifs suivants :

1. **Monitoring visuel** : la visualisation pourrait permettre d'avoir une vision globale d'une situation en temps réel afin de pouvoir repérer des événements qui se produisent. Dans le cadre du trafic, cela pourrait être des problèmes de congestion lors d'embouteillages. En effet, ces problèmes peuvent être enfouis dans la masse de données mise à disposition et seraient par conséquent compliqués à déceler sans une visualisation.
2. **Découverte de patterns** : la visualisation pourrait donner de nouvelles connaissances sur les données utilisées en permettant de dégager des patterns inconnus jusque là. En effet, la découverte de patterns est un enjeu crucial en mobilité car ceux-ci donnent des informations très importantes sur la mobilité et les caractéristiques des données concernées. Par exemple, la visualisation pourrait permettre de découvrir quelles routes sont empruntées par les automobilistes pour se rendre à leur travail.
3. **Clustering de patterns** : découvrir des patterns est intéressant mais cela le devient encore plus lorsque ces patterns peuvent être mis dans des *clusters* afin d'en tirer des généralités. En effet, lorsque des énormes masses de données sont utilisées, il peut être important d'agréger ces données dans des *clusters*. Cet objectif est donc lié au précédent. Pour repartir de l'exemple du point précédent, des *clusters* pourraient regrouper les routes en fonction de leur proximité géographique.

4. **Confirmation d'hypothèses** : les experts émettent des hypothèses concernant des aspects de mobilité sur lesquels ils travaillent. La visualisation peut aider à confirmer ou infirmer ces hypothèses.

## 2.4.2 Techniques de visualisation de trafic

Afin de définir les techniques de visualisation existantes, il est nécessaire de s'intéresser aux différents types de données que le trafic fournit. Il en existe principalement trois : les données temporelles, les données spatiales et les données spatio-temporelles (Chen *et al.*, 2015) (Andrienko *et al.*, 2017). Pour chacune d'entre elles, une visualisation peut être définie. Chaque type de donnée ainsi que les techniques de visualisation correspondantes sont explicités dans les paragraphes suivants.

**Visualisation du temps** La visualisation de données temporelles dépend de la manière dont le temps est défini. Celui-ci peut être linéaire, c'est-à-dire allant d'un début vers une fin. Il peut être également périodique, c'est-à-dire qu'il fait référence à un aspect récursif comme par exemple les saisons qui recommencent chaque année. Il peut enfin être à ramification, ce qui veut dire qu'il se concentre sur les interactions entre divers acteurs au cours du temps. Pour Chen *et al.* (2015), cette dernière définition du temps n'est pas encore utilisée dans des visualisations de données de trafic, elle n'est donc que mentionnée dans cet état de l'art. Par contre, les définitions de temps linéaires et périodiques sont bel et bien utilisées.

**Temps linéaire** Cette définition est la plus commune et la plus utilisée dans la représentation de données. Les principales techniques de visualisation utilisées dans ce cadre sont les *line charts* et les *stacked graphs*.

Les *line charts* ont l'avantage d'être faciles à lire lorsque l'objectif est de mesurer l'évolution d'une variable dans le temps. Un exemple est montré à la figure 2.3. Cette figure représente les pourboires par voyage donnés aux taxis originaires de différentes régions dans la période du 1<sup>er</sup> mai 2011 au 7 mai 2011 (Ferreira *et al.*, 2013). Cependant leur principal inconvénient est qu'ils ne peuvent pas facilement montrer des informations sur un grand nombre de variables différentes.

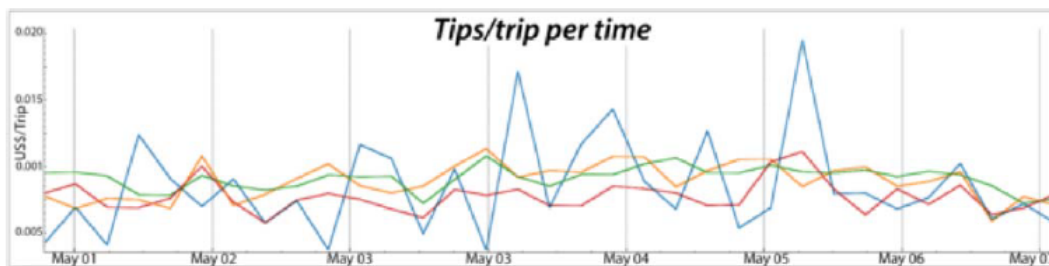


FIGURE 2.3 – Exemple d'utilisation de *line chart* pour représenter les pourboires par voyage donnés aux taxis originaires de différentes régions dans la période du 1<sup>er</sup> mai 2011 au 7 mai 2011 (d'après (Ferreira *et al.*, 2013)).

Les *stacked graphs* (Byron et Wattenberg, 2008) quant à eux permettent de visualiser l'évolution de plusieurs variables à travers le temps afin de les comparer. La figure 2.4(b) montre un exemple de *stacked-graph* visualisant le volume du trafic à une intersection de route (Guo *et al.*, 2011). Le temps linéaire peut être utilisé dans des visualisations de trafic pour montrer l'évolution de celui-ci à travers le temps en mettant en avant des pics de trafic et quand ils ont lieu.

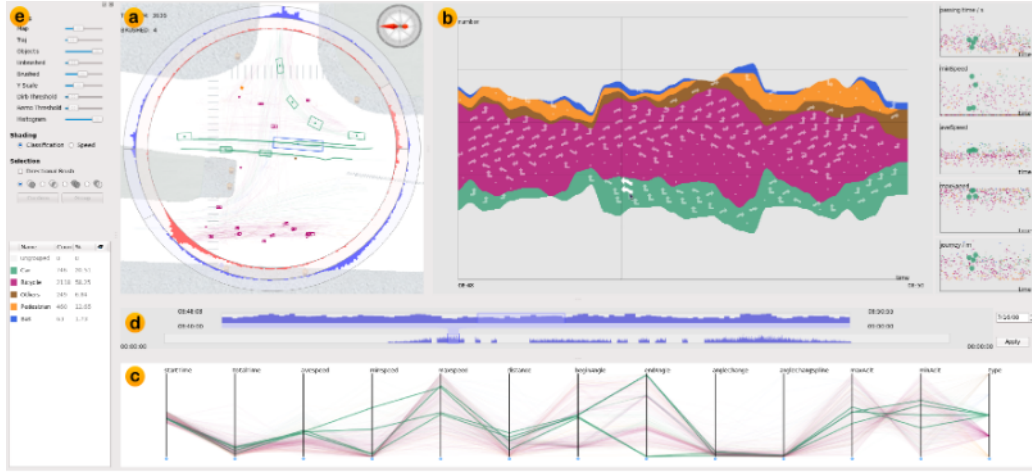


FIGURE 2.4 – Exemple d'utilisation de *stacked graph* (b) dans TripVista, outil permettant de trouver des *patterns* dans les flux de trafic aux intersections de routes. Les variables visualisées sont les directions des voitures et le volume de voitures (d'après (Guo et al., 2011)).

**Temps périodique** Chen et al. (2015) explique que la visualisation de phénomènes cycliques se fait principalement via un *radial layout*. Un *radial layout* est une technique de visualisation développée pour montrer de la récursion dans les données. Ces données sont le plus souvent représentées dans un graphe en forme circulaire, un *ring map*, comme dans la figure 2.5 qui est un exemple de représentation d'intensité de trafic au cours d'une journée (Pu et al., 2013). Chaque anneau représente un jour de la semaine. Ceci peut être comparé à une horloge.

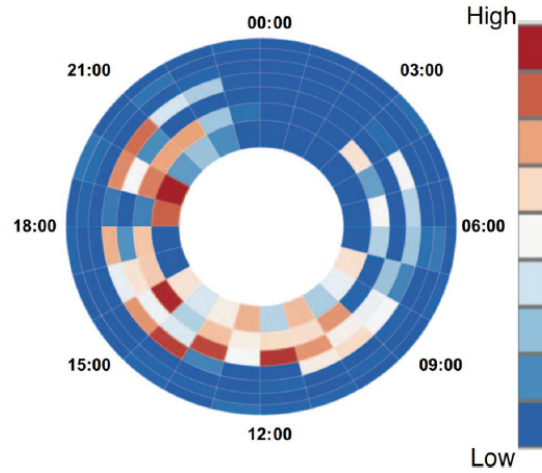


FIGURE 2.5 – Exemple de *ring map* basé sur un *radial layout* montrant la quantité de trafic par heure d'une journée durant une semaine (d'après (Pu et al., 2013)).

**Visualisation de l'espace** La notion d'espace est importante dans les données de trafic car elle permet de lier des événements. Ces événements peuvent être, par exemple, des accidents ou des embouteillages à une position géographique. Plusieurs catégories de techniques de visualisation existent : la *point-based*, qui utilise le symbole cartographique du point, la *line-based*, qui utilise le symbole cartographique de la ligne et enfin la *region-based*, qui utilise le symbole cartographique de l'aire. Ces catégories sont différenciées par leur niveau d'agrégation.

**Point-based** Cette catégorie part des positions discrètes des données et propose des techniques permettant de mettre en avant les points les plus intéressants. Cepen-



dant, les données continues ne peuvent pas être traitées. Celles-ci sont gérées par les techniques *line-based*. Les techniques *point-based* sont les suivantes :

1. Une visualisation ***dot-based*** positionne chaque point sur la carte. Cela permet à l'utilisateur d'avoir des informations très précises sur chaque donnée. Kloeckl et al. (2016) donnent un exemple de visualisation en temps réel de trains permettant de connaître leur position, le nombre de passagers ainsi que s'il a du retard ou non. Cependant lorsque le nombre de points devient très grand, la visualisation peut devenir confuse.

Dans ce cas, plusieurs solutions sont envisageables. Tout d'abord, il est possible d'utiliser des techniques d'interaction pour réduire cette confusion. Un exemple est la technique de la loupe. Cette technique ajoute une loupe sur la carte. Grâce à cette loupe, l'utilisateur peut zoomer sur la zone qui l'intéresse. Cela permet de réduire la confusion et de se concentrer sur quelques points. La figure 2.6 montre un exemple de loupe placée sur une carte permettant à l'utilisateur de se concentrer uniquement sur les points qui l'intéressent (Cruz, 2012). Une autre solution pourrait être de changer de technique de visualisation et d'utiliser une *heatmap*.

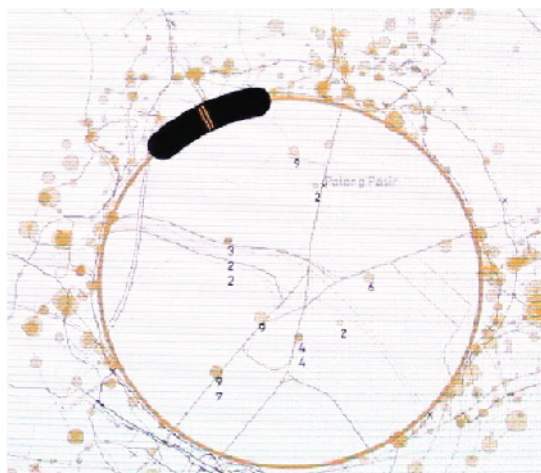


FIGURE 2.6 – Exemple de loupe placée sur une carte (d'après (Cruz, 2012)).

2. Une ***heatmap*** agrège les points et les représente sur une carte via des zones de différentes couleurs. Les couleurs donnent des informations sur le nombre de points qu'elle contient. La figure 2.7 donne un exemple de *heatmap* dans une ville (Liu et al., 2013). Plus la couleur d'une zone tend vers le rouge et plus elle contient des points.

Une *heatmap* est similaire à une *choropleth map* car toutes deux agrègent des données en fonction de zones géographiques. Cependant ces deux techniques ne doivent pas être confondues. Dans une *heatmap*, les données créent les zones en fonction de leur densité. Ces zones sont alors visualisées via des couleurs sur la carte. Dans une *choropleth map* par contre, qui est une technique *region-based*, les zones sont définies au préalable. Ces zones peuvent être par exemple des pays ou des régions. Les données sont ensuite agrégées dans ces zones et puis visualisées avec des couleurs (Torguson, 2016).



FIGURE 2.7 – *Heatmap* représentant le volume de trafic d’une ville (d’après (Liu et al., 2013)).

**Line-based** Les techniques *line-based* sont utilisées pour montrer des données continues telles que des trajectoires ou des flux.

Une trajectoire contient des données temporelles et spatiales. Les données spatiales donnent une position de la trajectoire à un instant donné tandis que les données temporelles donnent un historique du mouvement (Chen et al., 2015). Cette trajectoire est représentée sur une carte par une ligne droite ou courbe. Un flux représente une agrégation de mouvements entre différents lieux et est donc une agrégation de trajectoires. Cette agrégation peut se réaliser de deux manières différentes (Andrienko et Andrienko, 2013).

Soit les trajectoires sont agrégées par paires origine-destination (OD) sans prendre en compte les positions intermédiaires. Le flux a donc une origine, une destination et un nombre d’objets, magnitude dans la suite de ce mémoire, réalisant ce mouvement. Ces flux sont appelés flux OD dans la suite de ce mémoire. Un exemple d’utilisation de flux OD dans une visualisation est montré à la figure 2.11 (Stephen et Jenny, 2017). Chaque flèche représente un flux. Ces derniers sont des flux OD car les flèches donnent des informations sur l’origine, la destination et la magnitude des flux mais pas sur leurs positions intermédiaires.

Soit chaque trajectoire est représentée par un ensemble de positions intermédiaires entre l’origine et la destination. Dans ces conditions, le flux est l’agrégation de ces positions pour toutes les trajectoires. Ces flux sont appelés flux-transitions dans la suite de ce mémoire. La figure 2.8 présente un exemple de visualisation de trajectoires de navires agrégés dans des flux-transitions (Andrienko et Andrienko, 2013). Chaque flux est représenté par un ensemble de flèches qui donnent des informations sur ses positions successives. L’épaisseur d’une flèche représente la magnitude de la portion de flux correspondante.

Quelle que soit la manière d’agréger les trajectoires, la visualisation de flux est divisée en trois familles de techniques : *OD matrix*, *OD map* et *flow map* (Andrienko et Andrienko, 2013).

**OD matrix** Les données sont représentées dans une matrice. Les colonnes et les lignes correspondent respectivement à l’origine et la destination du flux tandis que la cellule donne une information sur la magnitude de ce flux. Les cellules de la matrice sont souvent colorées. Cette couleur permet de voir rapidement quels sont les flux ayant la plus grande magnitude.

Le principal inconvénient de cette méthode est qu’à part le nom des lieux, elle ne fournit pas d’information spatiale. La figure 2.9 montre un exemple d’*OD matrix* visualisant le nombre de relocalisations d’entreprises d’un état à un autre au sein des États-Unis (Guo et al., 2006).

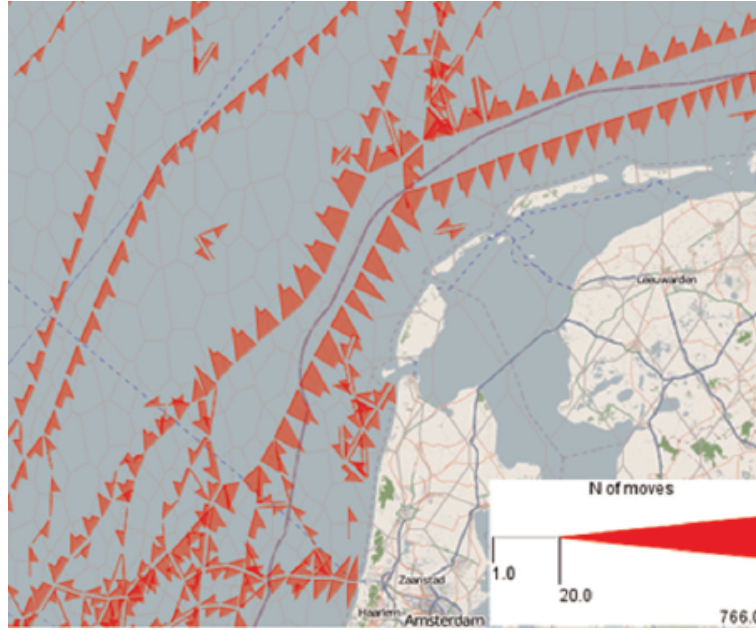


FIGURE 2.8 – Exemple de visualisation de trajectoires de navires agrégées dans des flux-transitions. Chaque flux est représenté par un ensemble de flèches qui donnent des informations sur ses positions successives. L'épaisseur d'une flèche représente la magnitude de la portion de flux correspondante (d'après (Andrienko et Andrienko, 2013)).

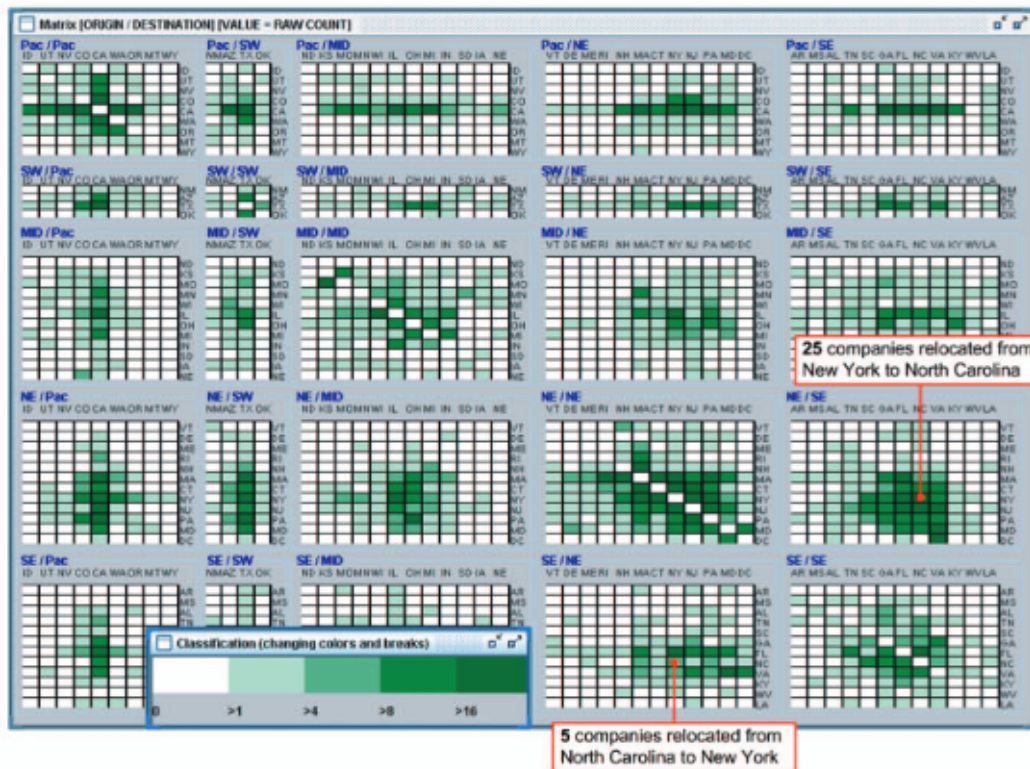


FIGURE 2.9 – Visualisation du nombre de relocalisations d'entreprises d'un état à un autre des États-Unis, via un *OD matrix* (d'après (Guo et al., 2006)).

**OD map** Cette technique se base sur une *OD matrix* pour visualiser les données. Chaque flux OD est représenté par une cellule d'une matrice. La colonne et la ligne de la cellule représentent respectivement l'origine et la destination du flux. Une *OD map* se différencie d'une *OD matrix* par le positionnement

des cellules. En effet, celles-ci sont organisées pour exprimer la position géographique de la donnée qu'elle contient (Wood *et al.*, 2010).

La figure 2.10 montre un exemple de visualisation d'un flux entre deux points situés en France via une *OD map* (Wood *et al.*, 2010). (a) représente un partitionnement de la France en de multiples cases identiques formant une grille, chacune représentant une portion géographique du pays. Pour arriver à (b), chaque case est divisée en autant de cases qu'il y en a dans la grille de base. Cette subdivision permet de représenter une trajectoire entre une origine et une destination par une seule cellule de la matrice présente à la figure (b). Par exemple la trajectoire rouge, qui a pour origine et destination respectivement (A,2) et (E,5) dans (a), sera représentée dans la cellule (AE,25) de (b). Par exemple la trajectoire inverse est visualisée dans la cellule (EA,52) de (b).

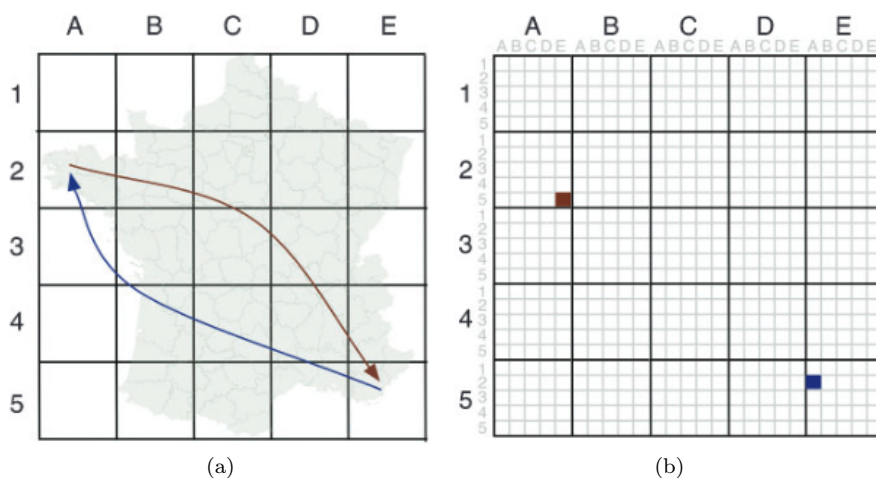


FIGURE 2.10 – Visualisation de deux trajectoires entre deux points situés en France ; (a) : la France est partitionnée en une grille de cellules ; (b) : chaque cellule est divisée en autant de cellules que dans (a). La trajectoire rouge de (A,2) vers (E,5) dans (a) est représentée par la cellule (AE,25) dans (b). La trajectoire inverse est visualisée dans la cellule (EA,52) (d'après (Wood *et al.*, 2010)).

**Flow map** Les flux sont représentés avec des lignes ou des flèches entre les différents lieux sur une carte. La magnitude du flux est représentée de différentes manières en fonction de la *flow map* générée. Elle peut notamment être symbolisée par une couleur. Par exemple, un flux tendant vers le rouge aura une magnitude plus élevée qu'un flux tendant vers le jaune. Elle peut aussi être représentée via l'épaisseur du symbole.

La figure 2.11 montre un exemple de *flow map* visualisant les 50 plus grands flux de migration entre les différents états des États-Unis entre 2009 et 2013 (Stephen et Jenny, 2017). Dans cette figure, chaque flèche représente un flux et son épaisseur visualise la magnitude de ce flux par rapport aux autres présents sur la *flow map*. Ces données utilisent une mesure de type *ratio* qui offre donc la possibilité d'effectuer des comparaisons poussées sur les flux. Les *flow map* sont discutées plus en détail dans la section 2.5.



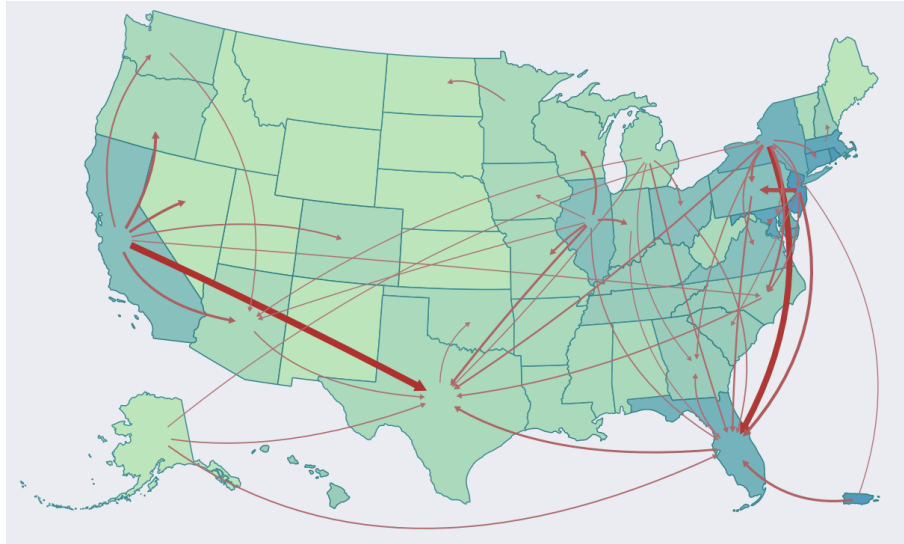


FIGURE 2.11 – Visualisation des 50 plus grands flux de migration entre les états des États-Unis de 2009 à 2013 via une *flow map*. Chaque flux de migration est représenté par une flèche. L'épaisseur de celle-ci représente la magnitude du flux (d'après (Stephen et Jenny, 2017)).

**Region-based** Les techniques *region-based* permettent de visualiser des situations en fonction de zones géographiques. Pour ce faire, les données sont agrégées dans des régions grâce à différentes règles. Les régions peuvent être définies par des entités administratives, comme dans la figure 2.12, ou par des polygones pour former un quadrillage. Par exemple, une ville pourrait être divisée en une multitude de carrés. Ces techniques permettent de révéler des *patterns* généraux dans les données mais sont inefficaces en ce qui concerne des *patterns* plus précis.

Ces techniques de visualisation peuvent intervenir dans deux situations. Les données disponibles peuvent tout d'abord avoir une sémantique qui implique une notion de région. Par exemple, une *choropleth map* est sémantiquement définie comme une technique *region-based* car les données sont agrégées dans des régions représentées via des couleurs sur la carte. Des techniques *region-based* peuvent aussi être utilisées lorsque le nombre de trajectoires ou de flux rend les techniques *line-based* inutilisables. La figure 2.12 montre un exemple de *choropleth map* et de *flow map* (Stephen et Jenny, 2017). La partie *choropleth map* de cette figure visualise la densité de population de chaque comté du Tennessee.

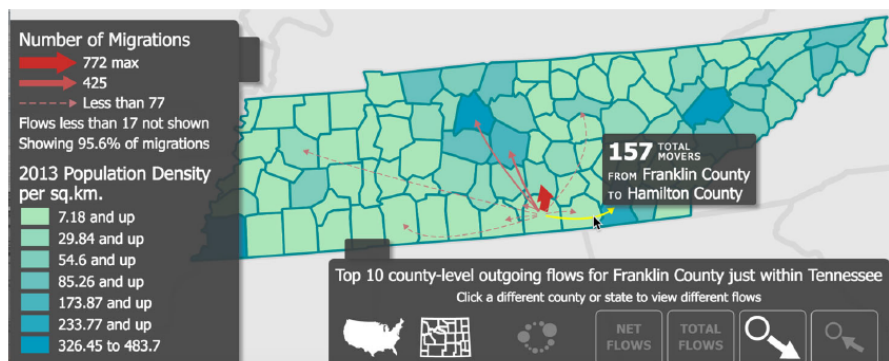


FIGURE 2.12 – Visualisation jumelant une *flow map* et une *choropleth map*. La *flow map* visualise les dix plus grands flux partant du comté de Franklin vers d'autres comtés du Tennessee. La *choropleth map* visualise la densité de population au kilomètre carré de chaque comté du Tennessee (d'après (Stephen et Jenny, 2017)).

**Visualisation de l'espace-temps** La visualisation de données spatio-temporelles amène un lien entre les visualisations temporelles et spatiales discutées ci-dessus. La technique de visualisation propre à l'espace-temps la plus connue est le *Space-Time-Cube*, STC dans la suite de ce mémoire (Kraak, 2003).

Un STC permet de visualiser des trajectoires dans un repère en trois dimensions. Les dimensions X et Y donnent des informations sur la position géographique des points de la trajectoire tandis que la dimension Z représente le temps écoulé. Cette technique est fort populaire (Andrienko et al., 2017) (Andrienko et al., 2016b) mais elle peut être compliquée à interpréter. En effet, la représentation d'une figure en trois dimensions sur un écran en deux dimensions implique un effort cognitif à fournir de la part de l'utilisateur (Ware, 2008) (St. John et al., 2001).

La figure 2.13 donne un exemple de visualisation d'un trajet d'une personne durant une journée ordinaire de travail (Kraak, 2003). Dans cette figure, le temps est représenté de bas en haut.

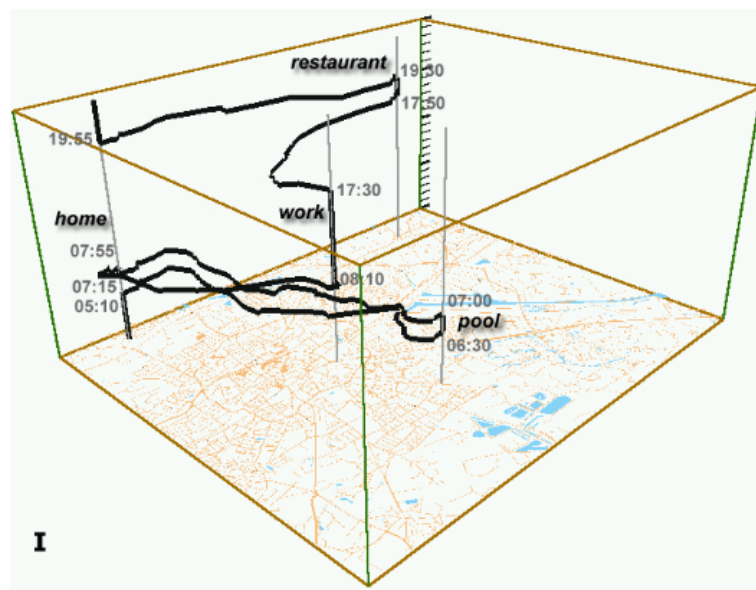


FIGURE 2.13 – Visualisation d'un trajet d'une personne durant une journée ordinaire de travail. Les axes X et Y représentent les coordonnées géographiques sur la carte tandis que l'axe Z représente le temps écoulé (d'après (Kraak, 2003)).

Cependant, le STC n'est pas la seule technique qui existe. Les techniques de visualisation de l'espace peuvent aussi être utilisées pour visualiser des données spatio-temporelles via des techniques d'interaction comme le *filtering*. Cette technique d'interaction permet de n'afficher que les données spatiales correspondant à une valeur temporelle. Par exemple, une *heatmap* telle que présentée à la figure 2.7 (Liu et al., 2013) pourrait être jumelée à du *filtering* pour montrer le volume du trafic d'une ville à une heure précise de la journée.

D'autres techniques d'interaction ont été définies et il est intéressant de les citer afin de montrer l'étendue des possibilités de visualisation qui existent pour répondre à une problématique. Gondim et al. (2014) décrivent deux projets qui se sont basés sur des métaphores afin de définir des visualisations et des techniques d'interaction. Le premier, *Walls Map Piece*, visualise la densité de trafic de rues avec des murs positionnés sur ces rues. Plus le trafic est dense et plus les murs sont grands. Le second, *Traffic Noise Mixer*, associe un son à la carte afin de montrer l'état du trafic d'une région via un autre sens que la vue. Plus le nombre de voitures d'une région ciblée est grand, plus le bruit entendu par l'utilisateur est fort.

**Visualisation de propriétés multiples** Les données temporelles et spatiales sont les plus importantes. Cependant les données peuvent contenir d'autres attributs. Ces autres propriétés peuvent être classées en trois catégories (Chen et al., 2015) :

1. Les **propriétés numériques** sont définies par des valeurs quantitatives. Elles décrivent un attribut numérique de la donnée comme par exemple une vitesse, une taille ou encore une aire. Ces propriétés peuvent changer avec le temps. Dès lors, des techniques de visualisation du temps peuvent être utilisées.
2. Les **propriétés catégoriques** définissent des variables décrivant l'état de la donnée. Cet état donne une information sur la catégorie de la donnée en question. Ces catégories peuvent être visualisées via des couleurs. Par exemple, des données sur des accidents de voiture peuvent être catégorisées par le type de véhicule, le type d'accident ou encore la direction du véhicule au moment de l'accident. Chaque couleur est assignée à une valeur qui représente une catégorie. Dans ces conditions, il est important de bien choisir les couleurs utilisées et le système *ColorBrewer* (Harrower et Brewer, 2003) apporte un élément de réponse à cette problématique.
3. Les **propriétés textuelles** donnent des informations supplémentaires sur les données de trafic sous forme de mots. Différentes techniques de visualisation peuvent être utilisées comme la visualisation *text-based* et en particulier la technique du nuage de mots (Rivadeneira et al., 2007) permettant de visualiser des mots en fonction de leur importance. D'autres techniques de visualisation de propriétés textuelles existent et sont reprises dans (Kucher et Kerren, 2015).

La visualisation simultanée de propriétés numériques et catégoriques peut principalement se faire de deux manières différentes : via une variation du STC et le *Parallel Coordinate Plot* (PCP).

**Variations du *Space Time Cube*** Le STC de base ne permet que de visualiser des données spatio-temporelles. Afin d'ajouter des attributs supplémentaires à cette technique, des variantes ont été définies comme par exemple le *GeoTime* qui affiche ces attributs sur la carte au même titre que les trajectoires (Kapler et Wright, 2005) ou encore le *stacking-based* STC qui permet de visualiser des trajectoires empilées les unes sur les autres sur un STC (Tominski et al., 2012). La figure 2.14 visualise, avec un *stacking-based* STC, l'état de congestion d'une autoroute de San Francisco entre 14h et 17h entre le 17 mai et le 10 juin 2008 (Tominski et al., 2012).

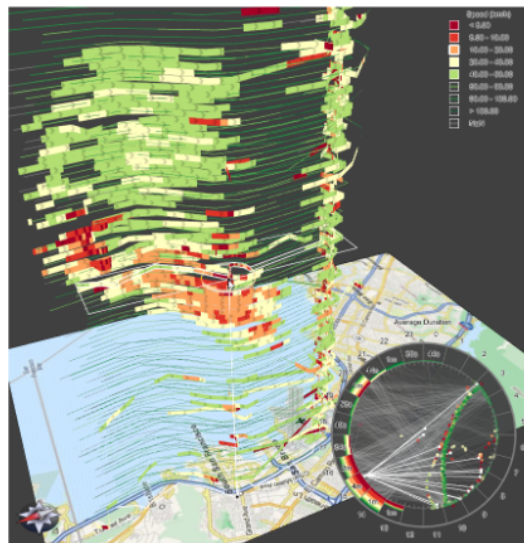


FIGURE 2.14 – Visualisation de la congestion sur une autoroute à San Francisco entre 14h et 17h entre le 17 mai et le 10 juin 2008 via un *stacking-based* STC. Plus les segments sont rouges et plus l'autoroute est congestionnée (d'après (Tominski et al., 2012)).

**Parallel Coordinate Plot** Un PCP est un graphique contenant de multiples axes parallèles les uns aux autres, chaque axe représentant un attribut. Une donnée est représentée par une ligne passant d'axe en axe et touchant l'axe à un endroit dépendant de la valeur pour l'attribut correspondant (Chen et al., 2015). Cette technique de visualisation permet de traiter des données possédant un grand nombre d'attributs. La figure 2.15 montre un exemple de visualisation d'accidents de la route utilisant le PCP (Pack et al., 2009).

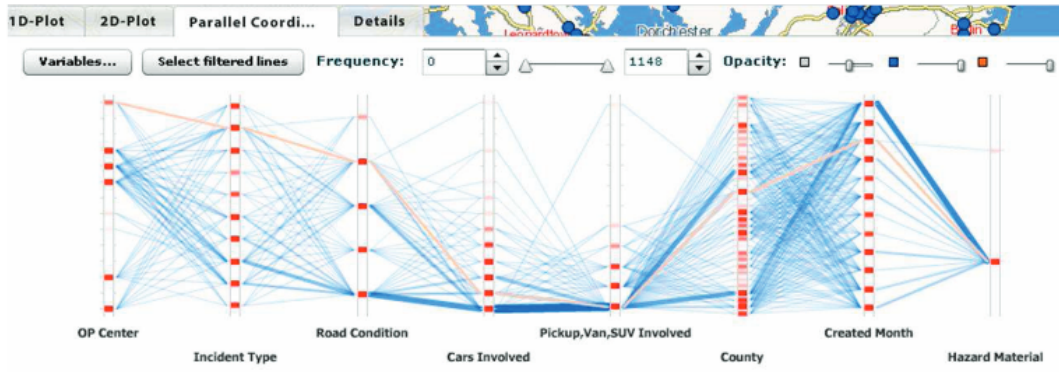


FIGURE 2.15 – Visualisation de données sur des incidents de trafic via un PCP. Chaque axe vertical représente un attribut qui peut être numérique, catégorique, spatial ou temporel. Chaque donnée est un accident de la route et est représenté par une ligne qui passe d'axe en axe (d'après (Pack et al., 2009)).

### 2.4.3 Croisement des objectifs et des techniques de visualisation

La table 2.1 présente un croisement effectué entre les objectifs de mobilité définis dans la section 2.4.1 et les techniques de visualisation de trafic définies dans la section 2.4.2.

Seuls trois des quatre objectifs sont dans cette table car la confirmation d'hypothèses peut être liée à n'importe quelle de ces techniques de visualisation. En effet celles-ci sont implémentées pour aider les experts à acquérir de l'information sur une situation de mobilité sur laquelle ils travaillent. Dans certains cas, ces experts auront déjà défini des hypothèses qui seront confirmées ou infirmées par les visualisations, lesquelles seront définies par l'une ou l'autre technique.

Chaque cellule de la table 2.1 contient les techniques provenant d'une des familles de techniques de visualisations (colonne) qui peuvent être utilisées pour répondre à un objectif (ligne). Les cellules vides de la tables montrent que certaines techniques de visualisation ne peuvent pas répondre à certains objectifs. Chaque objectif et ses techniques correspondantes sont discutés ci-dessous.

Il est important de noter qu'en théorie, presque toutes les techniques de visualisation pourraient être utilisées pour chacun des objectifs. Seulement en pratique, l'utilisation de ces techniques est parfois pertinente et parfois non pertinente car elles ne se valent pas toutes en toute circonstance. Par exemple, certaines sont définies pour visualiser des objets précis, d'autres pour visualiser des événements plus globaux. Dès lors, l'objectif de visualisation influence beaucoup le choix d'utilisation de telle ou telle technique. La table 2.1 représente cette analyse pratique de l'utilisation de techniques de visualisation en fonction de l'objectif de visualisation de la mobilité poursuivi.

**Monitoring visuel** Différentes techniques de visualisation peuvent être utilisées pour avoir une vision globale d'une situation de mobilité en temps réel. Les techniques de visualisation du temps, *line chart*, *stacked graph* et *radial layout*, en font partie. En effet, un *monitoring* en temps réel peut requérir des visualisations de l'évolution temporelle de variables, que ce soit un temps linéaire ou périodique.



En ce qui concerne la visualisation de l'espace, des techniques *point-based* et *line-based* peuvent être utilisées. Il peut être intéressant d'avoir une vision de la position d'objets précis, via une visualisation *dot-based*, ou l'agrégation d'objets, via une *heatmap*. Un exemple fictif pourrait être une visualisation en temps réel de la position de bus dans une ville via la technique *dot-based* ou une *heatmap*. De la même manière, une *flow map* est pertinente. Un exemple d'utilisation d'une *flow map* et d'une *heatmap* est la visualisation du trafic en temps réel proposée par Google Maps. Une *choropleth map*, *OD matrix* ou encore *OD map* pourraient théoriquement être utilisées pour réaliser du *monitoring*. En pratique, ces trois techniques de visualisation sont surtout utiles pour représenter des données historiques. C'est pourquoi elles ne sont pas présentes pour l'objectif de *monitoring* dans ce tableau.

Le STC simple n'est pas pertinent dans du *monitoring*. En effet, cette technique de visualisation de l'espace-temps n'est intéressante que lorsque le nombre de données est très limité or le *monitoring*, dans le cadre d'une ville, peut impliquer un très grand nombre de données à surveiller. Dans de telles circonstances, l'utilisation d'un STC rendrait la visualisation chaotique.

Cependant des variations du STC peuvent être utilisées. Fang et Lu (2011) discutent de l'implémentation d'une variation du STC pour visualiser l'évolution presque en temps réel de la pollution urbaine à Houston. En effet, les données ne sont récupérées que toutes les heures. Ce n'est donc pas du vrai temps réel. Néanmoins, Fang et Lu (2011) expliquent qu'il serait possible d'utiliser le STC avec des données en temps réel. Le PCP n'est pas pertinent dans ce cas de figure. En effet, il permet difficilement de visualiser une évolution temporelle dans les données. Il serait possible de l'utiliser si chaque axe représentait une valeur temporelle et qu'une donnée, en passant d'axe en axe, avançait dans le temps. Cependant cela reviendrait à implémenter un *line chart*.

**Découverte de *patterns*** En ce qui concerne la découverte de *patterns*, des techniques de visualisation du temps peuvent être utilisées. Les *line charts*, *stacked graphs* et *radial layouts* permettent tous de visualiser des *patterns*. En effet, le *line chart* et le *stacked graph* permettent de visualiser un certain nombre de données et donc de révéler d'éventuels *patterns* temporels. De même, un *radial layout* visualise de la récursion dans des données. Il peut tout à fait être utilisé pour découvrir des *patterns*.

Parmi les techniques de visualisation de l'espace, il est possible d'utiliser la *heatmap*, l'*OD matrix*, l'*OD map*, la *flow map* et la *choropleth map*. En effet, chacune de ces techniques est définie pour découvrir des *patterns*. La *heatmap*, jumelée à du *filtering*, peut révéler divers *patterns* dans les données. Les techniques *line-based OD matrix*, *OD map* et *flow map* sont définies pour révéler des *patterns* dans des trajectoires. Enfin, la technique *region-based choropleth map* agrège les données dans des régions et peut être utilisée pour découvrir des *patterns* globaux dans les données.

Pour implémenter une visualisation de l'espace-temps, le STC peut être utilisé. Cette technique peut révéler des *patterns* précis pour quelques données. Cependant lorsque le nombre de données est très grand, il serait judicieux de privilégier une technique de visualisation de l'espace, discutée dans le paragraphe précédent, avec une technique d'interaction comme le *filtering*.

Enfin, des techniques de visualisation de propriétés multiples peuvent être utilisées. Des variations du STC peuvent mettre en avant des *patterns* dans les données. Par exemple, la figure 2.14, qui utilise la variation *stacking-based* du STC, visualise la congestion sur une autoroute à San Francisco (Tominski et al., 2012). Cette visualisation révèle notamment les heures pendant lesquelles la congestion est la plus forte, en fonction de la portion d'autoroute concernée.

Un PCP peut aussi être utilisé afin de comparer les données en fonction de leurs attributs. De cette manière, des *patterns* peuvent être découverts en fonction de la position des données sur les axes. Par exemple, la figure 2.15 visualise des données sur

des incidents de trafic via un PCP (Pack et al., 2009). Sur cette visualisation, plus des données passent par le même endroit et plus la ligne résultante est épaisse. Ces lignes épaisses montrent des *patterns* qui semblent se répéter.

**Clustering de *patterns*** Pour effectuer un *clustering*, des algorithmes spécifiques doivent être utilisés. Il en existe un certain nombre et les plus importants, dans le cadre de ce mémoire, sont définis dans la section 2.5.1. Pour pouvoir réaliser ce *clustering* de *patterns*, les techniques de visualisation discutées dans cet état de l’art et les techniques de *clustering* doivent être utilisées de concert. Dans ces conditions, la plupart des techniques de visualisation peuvent être implémentées pour répondre à ce dernier objectif de la table 2.1.

Les techniques les moins pertinentes sont les suivantes. Une visualisation *dot-based* pourrait théoriquement être utilisée pour représenter différents *clusters*. Cependant cette technique est avant tout utilisée pour représenter l’état d’objets précis et non des *clusters* macro tels que des *clusters* de *patterns*. Les *OD matrix* et *OD maps* sont définis pour représenter des flux et non des *clusters*, contrairement à une *flow map* qui peut représenter des *clusters* de flux via des techniques de réduction de confusion comme l’*edge bundling* ou la simplification spatiale. Enfin, le STC n’est pas non plus pertinent car il est défini pour visualiser l’évolution d’une trajectoire à travers le temps. De ce fait, il est compliqué d’y insérer la notion de *cluster*.

TABLE 2.1 – Tableau croisant les objectifs de mobilité définis dans la section 2.4.1 avec les techniques de visualisation de trafic définies dans la section 2.4.2.

Techniques de visualisation						
	Visualisation du temps		Visualisation de l'espace			Visualisation de l'espace-temps
	Temps linéaire	Temps périodique	Point-based	Line-based	Region-based	
Objectifs visualisation mobilité	Monitoring visuel Découverte de patterns Clustering de patterns	Line chart Stacked graph Radial layout	Dot-based Heatmap	Flow map	/	Variations du STC
			Heatmap	OD matrix OD map Flow map	Choropleth map	Variations du STC PCP
				Flow map	/	

## 2.5 Zoom sur les *flow maps*

Comme discuté dans la section 2.4.2, une *flow map* est utilisée afin de visualiser des données de flux OD sur une carte. Elle permet de révéler des patterns macro quant à la position, la direction ou la magnitude des flux qu'elle visualise. Elle donne un point de vue de haut-niveau de l'ensemble des données qu'elle traite. De ce fait, elle n'est pas définie pour une analyse micro des données.

Afin d'être lisible, les flux représentés doivent éviter de se croiser ou de se recouvrir. Guo (2009) explique que les *flow maps* traditionnelles sont utiles pour des petits *datasets*. Lorsque le nombre de flux devient trop grand, des problèmes d'occlusion peuvent arriver, ce qui a pour conséquence une certaine confusion.

Par exemple, la figure 2.16 montre un exemple de *flow map* visualisant la migration entre 48 états des États-Unis (Guo, 2009). Cette *flow map* n'est pas facilement interprétable car le nombre de flux représentés empêche une vision claire de la situation. Dans cet exemple, il est compliqué de récupérer de l'information sur la migration entre les 48 états et presque impossible d'espérer en récupérer au niveau de chaque comté de ces états.



FIGURE 2.16 – *Flow map* représentant la migration entre 48 états des États-Unis. Plus les flux sont importants et plus ils sont épais et ont une couleur tendant vers le rouge. Le grand nombre de flux représentés rend la visualisation moins efficace et plus confuse (d'après (Guo, 2009)).

Cette situation de confusion, lorsque le nombre de données est trop grand, amène un dilemme. Comment réduire cette confusion tout en gardant un maximum d'information dans la visualisation ? De nombreuses techniques tentent d'apporter leur éclairage sur la question, chacune ayant des avantages et des inconvénients. Elles doivent être utilisées au cas par cas car il n'existe pas de réponse universelle à ce dilemme.

Les sections 2.5.1 et 2.5.2 présentent deux techniques permettant de réduire la confusion, la première en agissant sur la définition de régions, la deuxième en agissant sur le nombre de flux visualisés. Enfin, la section 2.5.3 présente brièvement d'autres techniques de réduction de confusion. Toutes ces techniques ne sont pas mutuellement exclusives. Il n'est pas rare d'en utiliser plusieurs dans une même visualisation, pour autant qu'elles fassent sens.

### 2.5.1 Simplification spatiale

La simplification spatiale a pour objectif de réduire le nombre de flux en diminuant le nombre d'emplacements d'origine et de destination des flux, ceux-ci étant des points. Pour ce faire, ces emplacements sont agrégés dans des *clusters*. Leurs flux sont aussi agrégés afin de générer des flux entre *clusters* plutôt qu'entre points (Andrienko et al., 2016a). Cette technique réduit le nombre de flux présents sur la visualisation. Il existe plusieurs techniques de génération de *clusters* géographiques. En général, elles sont classées dans deux groupes : le *partitional clustering* et le *hierarchical clustering* (Guo et al., 2012).

**Partitional clustering** Cette famille d’algorithmes consiste à classer les données dans des *clusters* distincts. L’algorithme le plus connu est le *k-means* qui consiste à répartir les données dans  $k$  *clusters*,  $k$  défini au préalable, de manière à minimiser une certaine fonction. Dans le cas de la simplification spatiale, cette fonction à minimiser est la distance séparant le centre du cluster de tous ses points (Likas et al., 2003).

Les *clusters* peuvent aussi être générés via un algorithme *density-based*. Il définit les *clusters* comme des zones de haute densité de données. Contrairement à *k-means*, le nombre de *clusters* n’est pas connu au préalable et il peut également mettre en avant les *outliers*, c’est-à-dire les données ne se retrouvant dans aucun *cluster* (Kriegel et al., 2011).

**Hierarchical clustering** Cette famille d’algorithme consiste à placer les données dans des *clusters* hiérarchiques, c’est-à-dire plusieurs couches de *clusters* imbriquées les unes dans les autres (Guo et al., 2012). Par exemple, Guo (2009) utilise un algorithme de *hierarchical clustering* afin de réduire la confusion de la figure 2.16. Les origines et destinations des flux sont regroupés en régions de telle sorte qu’il existe plus de flux de migration internes à chaque région que de flux entre régions.

La figure 2.17 montre 45 régions définies de telle sorte que chaque région ne peut contenir moins de 3.5 millions d’habitants et doit maximiser la modularité c’est-à-dire une mesure permettant de comparer la migration entre chaque état en prenant en compte leur population (Guo, 2009). Ces *clusters* sont hiérarchiques car ils sont définis en plusieurs niveaux. Chaque *cluster* présent sur la figure 2.17 est une agrégation de plusieurs *clusters* plus petits qui représentent chacun un comté des États-Unis.

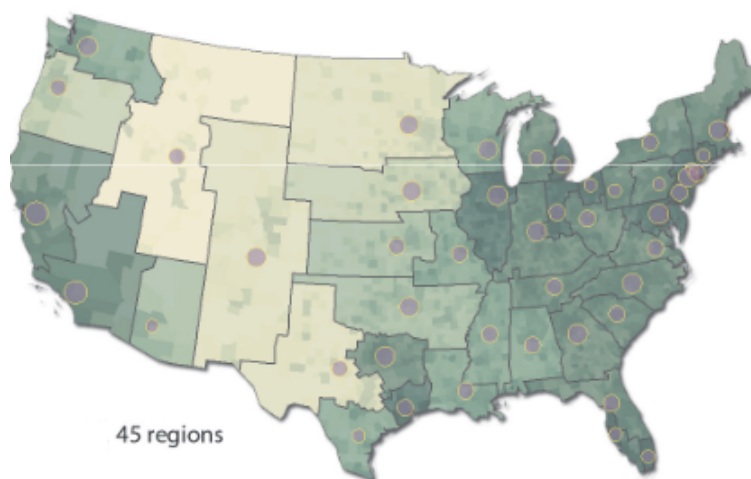


FIGURE 2.17 – Visualisation des 45 régions découvertes avec la méthode de régionalisation utilisée dans (Guo, 2009). Chaque région doit avoir un minimum de 3.5 millions d’habitants et maximiser ses connexions internes. Chaque cercle est positionné au centroïde défini par rapport à la population et représente la modularité de la région, c’est-à-dire la mesure permettant de comparer la migration entre chaque état en prenant en compte leur population. Le niveau de densité est représenté par les couleurs des comtés (d’après (Guo, 2009)).

## 2.5.2 Edge bundling

La technique de l’*edge bundling* consiste à regrouper les flux similaires dans des *bundles*, c’est-à-dire des paquets (Phan et al., 2005) (Buchin et al., 2011). Cette similarité peut être définie par la distance les séparant ou encore la similarité de l’origine ou de la destination du flux. Cela permet de réduire la confusion due au nombre de flux. Cette technique est définie pour les flux OD. Les flux-transitions utilisent déjà en quelque sorte l’*edge bundling* car les trajectoires sont rassemblées en flux en fonction de leur similarité, définie par leur position. Plusieurs variantes de cette technique existent (Zhou et al., 2013).

Les plus connues sont la *geometry-based* et la *skeleton-based*. La *geometry-based* fonctionne comme suit. Lors du *clustering*, elle impose aux flux de passer par certains endroits (Cui et al., 2008). La *skeleton-based*, de son côté, combine du *clustering* et du calcul de distances afin de construire itérativement les *bundles*. Les flux sont rassemblés dans ces *bundles* en fonction de ce calcul de distances (Ersoy et al., 2011).

Un exemple d'utilisation de ces approches est montré à la figure 2.18 visualisant les flux d'immigration aux États-Unis de trois manières différentes. La partie (a) présente les flux bruts sans utilisation d'*edge bundling* (Cui et al., 2008). Les parties (b) et (c) présentent ces mêmes flux visualisés respectivement via une approche *geometry-based* et *skeleton-based* de la méthode *edge bundling* (Cui et al., 2008) (Ersoy et al., 2011). Les deux dernières visualisations sont lisibles contrairement à la première dans laquelle le nombre de flux est tellement grand que presque tous les flux se recouvrent les uns les autres et empêche de retirer de l'information de la visualisation.

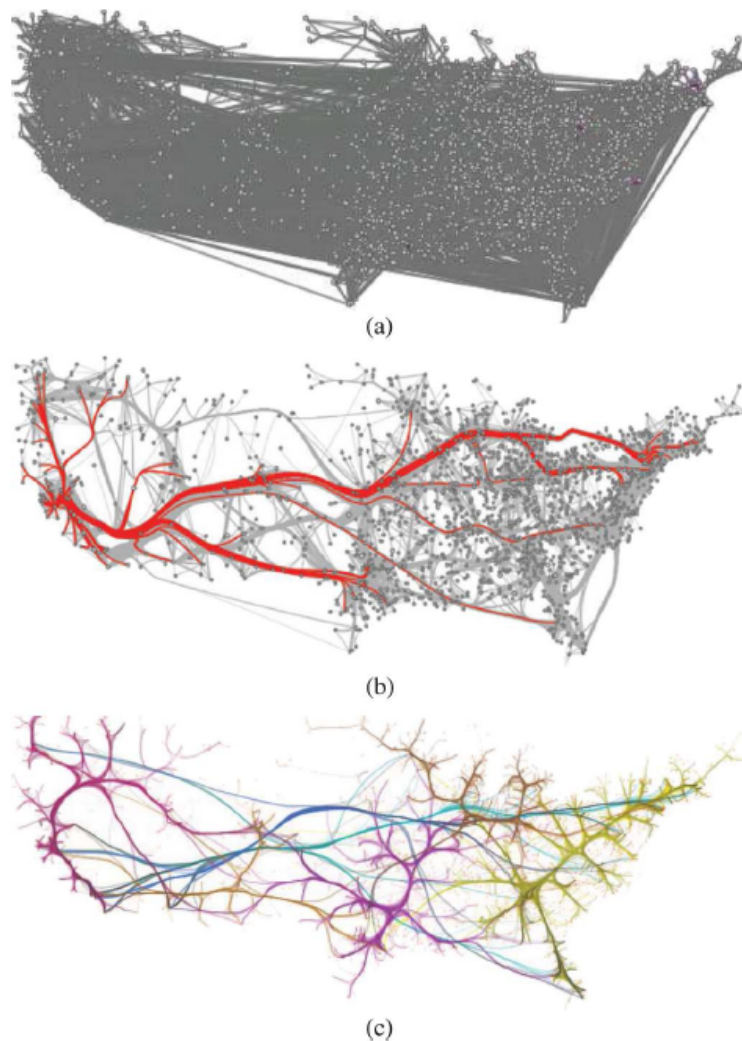


FIGURE 2.18 – (a) Visualisation *line-based* montrant les flux d'immigration aux États-Unis (d'après (Cui et al., 2008)); (b) Résultat d'un *edge bundling* généré à partir de (a) via une approche *geometry-based* (d'après (Cui et al., 2008)); (c) Résultat d'un *edge bundling* généré à partir de (a) via une approche *skeleton-based* (d'après (Ersoy et al., 2011)).

Cette technique de visualisation a tout de même quelques inconvénients. Tout d'abord, son utilisation sur une carte géographique pour des flux OD peut donner l'impression à l'utilisateur qu'il existe des routes alors que ce n'est pas le cas (Andrienko et Andrienko,

2013). En effet ces routes définies par l’algorithme sont des représentations des flux mais ne représentent pas forcément leurs positions géographiques réelles. Ce problème ne se pose pas pour les flux-transitions.

Ensuite, ces flux ne sont pas représentés via des flèches. Il peut donc être plus compliqué pour l’utilisateur de voir dans quelle direction vont les flux (Chen *et al.*, 2015). Guo et Zhu (2014) présente une technique n’utilisant pas l’*edge bundling* tout en permettant de visualiser la direction des flux. Elle consiste à estimer la densité de flux pour chaque paire OD et à ne visualiser que les flux les plus importants. Ersoy *et al.* (2011) explique que l’*edge bundling* n’est efficace que pour visualiser des flux venant ou allant à un ou deux endroits maximum. Par exemple, la figure 2.18 (b) visualise l’immigration en provenance ou à destination de l’ouest des États-Unis (Cui *et al.*, 2008).

### 2.5.3 Autres techniques de réduction de confusion

Outre la simplification spatiale et l’*edge bundling*, il existe d’autres techniques utilisées afin de réduire la confusion d’une *flow map*.

**Filtering** Cette technique consiste à ne visualiser que les données qui correspondent à un certain filtre. Elle est très souvent utilisée dans des visualisations. Le *filtering* permet ainsi de n’afficher que les données jugées intéressantes. Un exemple de filtre pourrait être de ne montrer que les flux ayant une magnitude plus grande ou plus petite qu’un certain seuil (Guo, 2009) ou encore uniquement les flux présents entre des positions sélectionnées (Van den Elzen et Van Wijk, 2014).

Elle peut être implémentée pour permettre à l’utilisateur de changer lui-même le filtre via une interface lui permettant d’interagir avec la visualisation (Wood *et al.*, 2011). Cette technique est utilisée dans la figure 2.11 pour ne montrer que les 50 plus grands flux de migration aux États-Unis entre 2009 et 2013 (Stephen et Jenny, 2017). Comme pour l’*edge bundling*, le *filtering* permet de réduire la confusion mais le coût de cette réduction est une perte d’information qui peut être importante suivant le filtre utilisé (Andrienko et Andrienko, 2013).

**Simplification visuelle** Cette technique consiste à retirer des éléments des flux sur la visualisation afin de diminuer la quantité d’informations présentes sur la *flow map*. Une première manière de procéder est de réduire l’opacité des flux ayant une petite magnitude (Wood *et al.*, 2011). De ce fait, les flux les plus importants seront plus visibles que les autres. Cependant cette technique implique une perte d’informations car les plus petits flux ne seront plus visualisés.

Une deuxième méthode consiste à supprimer partiellement ou totalement le milieu des lignes représentant les flux. Les origines sont représentées par une couleur et les destinations par une autre. Cela permet de donner une information sur le sens du flux. Cependant les flux sont moins faciles à visualiser car ils sont coupés en deux. La figure 2.19 montre un exemple de *flow map* visualisant les flux de réfugiés à travers le monde en 2008 (Boyandin *et al.*, 2010). Le milieu de chaque flux est presque invisible, son origine est représentée en vert et sa destination en rouge.



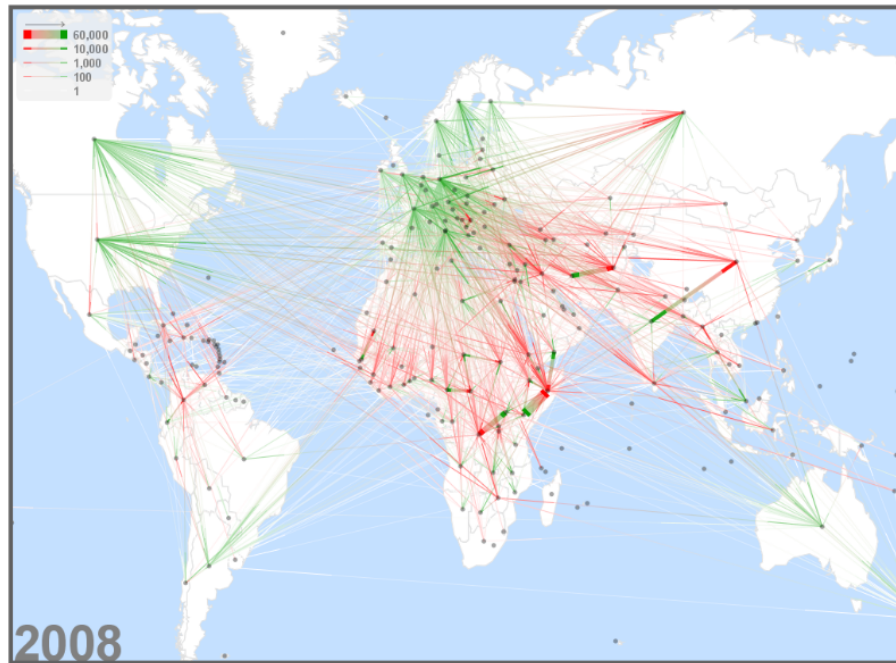


FIGURE 2.19 – Visualisation des flux de réfugiés dans le monde en 2008. Le milieu de chaque flux est presque invisible. Les origines sont représentées en rouge et les destinations en vert (d'après (Boyandin *et al.*, 2010)).

**Kernel Density Estimation (KDE)** Cette approche statistique consiste à calculer la densité des données afin de les montrer sur la carte via une *heatmap*. Plus il y a de flux à un endroit et plus celui-ci sera sombre sur la carte. La figure 2.20 montre un exemple d'utilisation de KDE pour visualiser la densité du trafic aérien aux États-Unis (Lampe et Hauser, 2011).

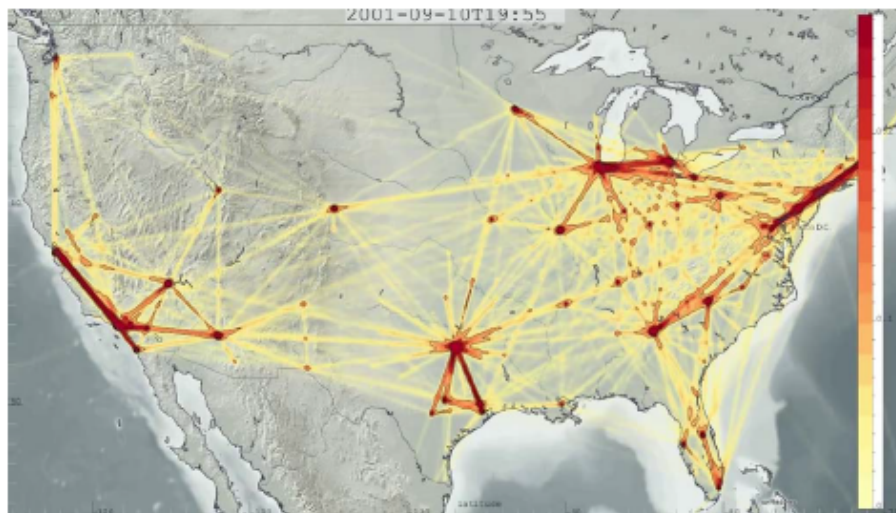


FIGURE 2.20 – Visualisation de la densité du trafic aérien aux États-Unis via une approche KDE. Plus la densité de flux est élevée à un endroit sur la carte et plus cet endroit sera de couleur sombre (d'après (Lampe et Hauser, 2011)).



## 2.6 Conclusion

Cet état de l'art rassemble de nombreuses techniques de visualisation de trafic, structurées d'une manière qui croise les types de données visualisées et les objectifs de visualisation. Ce croisement est réalisé dans la table 2.1. Cependant, d'autres techniques de visualisation peuvent exister en dehors de ce qui est couvert par l'état de l'art. Cette table pourrait donc être complétée par d'autres chercheurs. Elle pourrait aussi leur servir de base pour la réalisation de nouvelles classifications ou de *frameworks* de recommandation de techniques de visualisation de trafic.

La synthèse des techniques de visualisation de trafic a deux implications pratiques. Tout d'abord, cela permet de développer les chapitres suivants de ce mémoire et donc le développement d'une solution pour répondre à la question de recherche. Ensuite, cela donne des outils à la ville de Charleroi pour s'inscrire dans une mobilité intelligente. En effet, la table 2.1 est utile pour déterminer quelles techniques de visualisation utiliser pour répondre à un objectif de visualisation au niveau de la mobilité. Elle pourrait aider la ville de Charleroi à développer des solutions informatiques à base de visualisation pour répondre à une future problématique de mobilité.

## Chapitre 3

# Développement de la solution

La Ville possède un certain nombre de données liées à la mobilité, que ce soit des données statiques, comme des cadastres, ou des données dynamiques, comme des données sur la circulation automobile au jour le jour. Cependant, ces données brutes ne deviennent intéressantes que si elles permettent de répondre à une ou plusieurs problématiques de mobilité de la Ville. La difficulté est de trouver un lien entre ces données et les tâches de mobilité dont la Ville aimerait s'occuper. Dans ce contexte, le stage, d'une durée de trois mois, s'est déroulé à l'hôtel de ville de Charleroi dans le cabinet mobilité. Son objectif était de construire ce lien via une analyse des sources de données disponibles et le développement d'un outil de visualisation permettant d'aider les experts à prendre des décisions au niveau de la mobilité.

La méthodologie suivie tout au long du processus de développement de la solution est expliquée dans la section 3.1. Il est important de définir les données disponibles et les besoins de la Ville pour enfin définir une visualisation répondant à un de ces besoins. La section 3.2 présente un ensemble non exhaustif des données de mobilité dont la Ville dispose. La section 3.3 présente les besoins en matière de mobilité découverts lors des premières semaines du stage. La section 3.4 présente une solution proposée à un des besoins identifiés dans la section 3.3, choisi comme sujet du stage en concertation avec les clients.

### 3.1 Méthodologie

De manière générale, de nombreuses réunions ont été organisées avec différents organes de la Ville. Les premières réunions avaient pour but de présenter les données disponibles ainsi que le fonctionnement des différents organes de la Ville et les problématiques de mobilité auxquelles ces organes font face. Ces réunions ont permis de rassembler l'ensemble des données pertinentes qui sont présentées dans la section 3.2 et les besoins présentés dans la section 3.3.

Le développement d'un outil informatique implique des clients. Tout au long du processus, le client était la Ville. La plupart du temps, celle-ci était représentée par trois personnes :

1. Pascaline Gilson, attachée Mobilité à la ville de Charleroi. Après un parcours de cinq ans dans le privé et dix ans dans une administration, elle travaille maintenant notamment pour conseiller les politiques et les études de géographie spécialisées dans les transports. Pascaline a été la principale interlocutrice durant le développement de la solution.
2. Géraldine Ansotte, responsable projets à l'administration de la direction générale de la ville de Charleroi.
3. Cathy Libois, conseillère en mobilité à l'administration de la ville de Charleroi.

Le fait que la Ville soit représentée par ces personnes était très intéressant. En effet, elles ont des fonctions différentes, ce qui implique différents points de vue sur le besoin ainsi que la solution proposée. De ce fait, la solution a fait en sorte de répondre au mieux au besoin via la prise en compte de ces différents points de vue. Dans la suite de ce mémoire, ces trois personnes représentent le client. Lorsque le sujet du stage a été décidé, plusieurs *mockups* papier ont été réalisés. Cette technique permet de proposer rapidement un début de solution

au client sans devoir programmer, ce qui limite le temps perdu lorsque la solution doit être modifiée. Quelques *mockups* sont présentés dans la section 3.4.5.

Peu de temps après avoir défini la problématique du stage et la question de recherche de ce mémoire, le gouvernement belge a imposé un confinement pour lutter contre la pandémie de Covid-19. Cela n’a pas empêché le développement itératif de la solution proposée. En effet, des réunions ont été organisées à distance toutes les une à deux semaines avec les clients. L’objectif de ces réunions était de faire une démonstration de ce qui a été développé depuis la dernière réunion et de laisser l’occasion au client de faire des remarques afin de les incorporer au mieux dans le projet, que ce soit des remarques au niveau de l’interface ou au niveau des fonctionnalités proposées. En parallèle à ces réunions avec le client, d’autres réunions ont été organisées avec le promoteur et l’assistant encadrant le mémoire afin de présenter l’avancement du développement et leur permettre de faire des remarques sur les visualisations et principalement sur l’interface développée pour l’utilisateur final.

Ces réunions avec d’un côté le client et de l’autre des experts en visualisation ont permis de prendre en compte le plus de points de vue différents afin de proposer la solution la plus efficace, pertinente et simple à utiliser pour les clients.

La méthodologie suivie tout au long du processus de développement des visualisations pour satisfaire au mieux les besoins de la Ville est issue du *framework* WHAT WHY HOW (Munzner, 2014). Il a été créé pour développer des visualisations en trois étapes. Le développeur doit tout d’abord analyser en profondeur les données à disposition, ce qui correspond au WHAT du *framework*. En effet, il est inutile de vouloir développer une visualisation s’il n’existe aucune donnée pour le faire. La deuxième étape, le WHY, consiste à répondre à cette question : pourquoi développer une visualisation ? En bref, quels sont les besoins du client qui pourraient nécessiter l’utilisation d’une ou plusieurs visualisations ? En effet, s’il existe des données mais que l’utilisateur n’a aucun besoin, la visualisation ne servira à rien. Enfin, une fois que les données et les besoins du client sont définis, le HOW entre en jeu. Cette dernière partie consiste à analyser comment la visualisation doit être développée afin de répondre au besoin défini dans le WHY avec les données définies dans le WHAT. Les trois étapes sont importantes. Si l’une d’entre elles n’est pas bien réalisée, l’objectif du développement de la visualisation ne sera pas rempli.

La structure de ce chapitre est basée sur ce *framework*. L’ensemble des données disponibles, le WHAT, est discuté dans la section 3.2. Les différents besoins de la Ville, le WHY, sont identifiés dans la section 3.3. Enfin, le développement de la visualisation répondant à un besoin choisi comme sujet du stage en concertation avec les clients, le HOW, est discuté dans la section 3.4. Vu que ce développement répond à une problématique pour des données précises, le *framework* n’est explicité que pour ces données et le besoin concerné. Concrètement, pour chaque visualisation présentée dans la section 3.4.5, un tableau résumant l’analyse WHAT WHY HOW réalisée pour arriver à cette visualisation est présenté.

## 3.2 Données dont la Ville dispose

Différentes sources de données sont disponibles. Ces sources sont très hétérogènes et ne sont pas centralisées. La Ville possède très peu de données de mobilité la concernant et doit donc demander des données à plusieurs acteurs externes. Les sections suivantes présentent un ensemble de données de mobilité dont la Ville dispose. Cet ensemble n’est pas exhaustif. En effet, il regroupe les données découvertes lors des premières semaines du stage et il est fort possible que d’autres sources existent.

### 3.2.1 Données du Système d’Information Géographique

La première source de données est le SIG (Système d’Information Géographique) de la Ville et qui contient la plupart de ses données de mobilité. Comme son nom l’indique, un SIG permet de rassembler, traiter et analyser des données géographiques (Maliene et al., 2011). Les données de la Ville sont principalement statiques : elles n’évoluent que très peu dans le temps. Ce sont par exemple des données sur l’emplacement des arrêts de bus dans

la commune de Charleroi, des parkings détenus par des privés comme des magasins et des parkings publics, des pistes cyclables, des espaces verts ou encore des sens uniques limités.

Ces données sont encodées dans un format particulier : le *shapefile*. Ce format contient les informations géographiques d'une donnée, que ce soit un point, une ligne ou un polygone (ESRI, 1998). Ce type de donnée est le plus utilisé dans le cadre d'un SIG et permet de lier un fichier CSV avec ses données géographiques. Par exemple, un fichier CSV reprend des informations sur des parkings (à qui ils appartiennent, combien de places ils contiennent) tandis que le *shapefile* lié le positionne sur la carte.

Ces *shapefiles* sont utilisables dans QGIS, un logiciel SIG *open source* qui permet notamment de visualiser des *shapefiles* sur des cartes et de réaliser une série d'opérations dessus comme ajouter ou supprimer des données ou avoir des informations sur une donnée particulière en fonction de son positionnement sur la carte.

### 3.2.2 Données d'analyse de trafic de FLOWCheck

La deuxième source de données est la plateforme FLOWcheck, un outil d'analyse de trafic que la Ville a acquis fin 2019. Cette plateforme web donne des données de circulation à partir du 1er janvier 2017 jusqu'au 31 décembre 2021 inclus. Les données sont accessibles jusqu'à la veille, c'est-à-dire qu'un utilisateur peut accéder à toutes les données à partir du 1er janvier 2017 jusqu'à la veille du jour auquel il utilise la plateforme. Ces données sont disponibles à la demande, c'est-à-dire que l'utilisateur doit demander à la plateforme de lui fournir les données pour un lieu et une période précise, suite à quoi FLOWcheck, après quelques calculs, lui donne les résultats de sa requête sous forme de graphiques et de *heatmap*. La plateforme permet aussi de télécharger les données brutes de la requête. Les résultats dépendent du type de requête entrée par l'utilisateur et sont présentés dans la table 3.1. Un aperçu de la plateforme est disponible sur la figure 3.1.

TABLE 3.1 – Différentes analyses proposées par FLOWcheck

Requête	Spécifications de l'utilisateur	Résultat
Analyse de route	Une route pour une période de l'année, des jours de la semaine et un intervalle d'heures de la journée	Temps de parcours de la route, diagrammes de quartiles présentant la proportion de voitures considérées comme lentes en fonction du lieu où elles se trouvent et de l'heure à laquelle elles y sont.
Analyse de segment	Un segment de route pour une période de l'année, des jours de la semaine et un intervalle d'heures de la journée	Heatmap présentant quels véhicules ont été détectés sur le segment sélectionné au cours de la période choisie. Pour ces véhicules, l'analyse vérifiera les autres segments où ils ont aussi été détectés. Chaque tronçon de route indique la proportion de ces véhicules qui ont été détectés au cours de la même période de temps. Si la détection est en dehors de la période de temps, le véhicule est écarté des résultats.

TABLE 3.1 – Différentes analyses proposées par FLOWcheck

Requête	Spécifications de l'utilisateur	Résultat
Analyse de carte historique	Une zone de la Ville pour une période de l'année, des jours de la semaine et un intervalle d'heures de la journée	Heatmap présentant la vitesse mesurée par rapport à la vitesse maximale autorisée. Une couleur verte dans un centre-ville représentera une vitesse différente d'une couleur verte sur l'autoroute. Cependant, la couleur verte signifie une circulation fluide sur les deux routes.

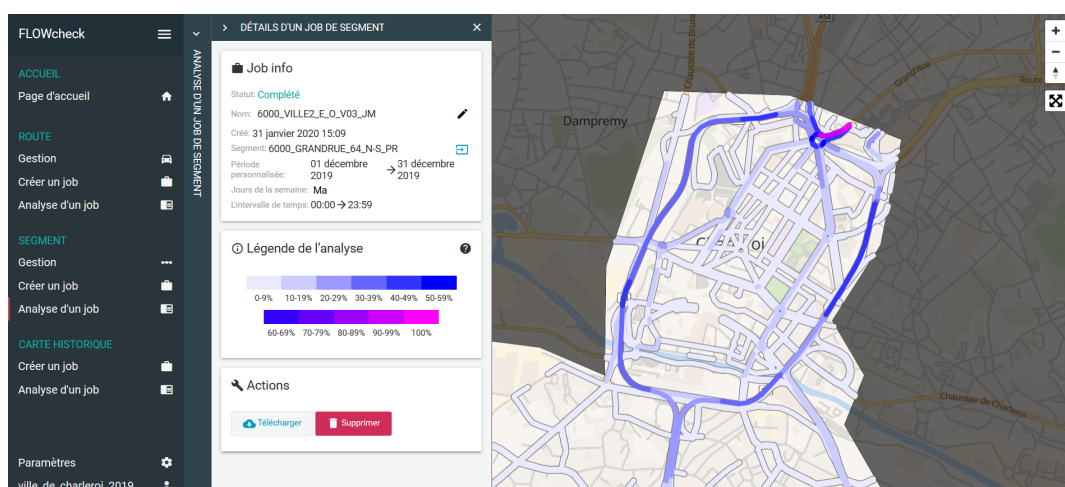


FIGURE 3.1 – Aperçu de l'analyse de segment de la plateforme FLOWcheck. Les différents menus ainsi que le détail de la visualisation sont présentés à gauche tandis que la visualisation en elle-même est à droite.

### 3.2.3 Données de déplacement de Proximus

En 2018, la Ville a acheté des données à Proximus concernant des déplacements de personnes entre plusieurs zones au sein de l'entité de Charleroi entre le 20 mars et le 21 mars 2018. Ces zones sont définies en fonction du positionnement des antennes Proximus et les données en fonction du nombre d'utilisateurs ayant un abonnement Proximus pour leur téléphone<sup>1</sup>. A partir de là, Proximus a pu tracer les utilisateurs de manière anonyme afin d'agrégier les données demandées par la Ville. Proximus a fourni un ensemble de visualisations montrant la répartition des déplacements au cours de la journée ainsi que le nombre de déplacements entre chaque zone en fonction d'un créneau horaire qui peut être défini par l'utilisateur. Les visualisations fournies sont des *flow maps* représentant ces flux entre zones au cours de la journée.

### 3.2.4 Données de la police de Charleroi

La police de Charleroi peut récupérer des données provenant de ses radars positionnés en ville. Ces données ont plusieurs attributs :

1. Selon Proximus, une grosse proportion des carolorégiens ont un abonnement chez eux. Les données récoltées sont donc, selon eux, pertinentes et représentatives.

- type d'infraction : qualification exacte de l'infraction avec un article de loi et une possibilité de les regrouper par thématique (par exemple vitesse, stationnement, défaut d'assurance ou de contrôle technique ou encore signalisation),
- type de constatation : avec ou sans interception,
- type de PV : perception immédiate ou PV classique,
- lieu et date de l'infraction,
- véhicules impliqués : type de véhicule, plaque, nationalité de la plaque,
- vitesse : autorisée, mesurée et corrigée,
- informations personnelles sur le contrevenant comme le nom, le prénom, la date de naissance ou encore l'adresse,
- montant de la perception et mode de paiement,
- type d'appareil radar utilisé

Ces données ne sont directement accessibles que depuis les ordinateurs de la police. Il est possible de leur demander d'envoyer des données spécifiques à une zone et une période précise. Cependant l'accès direct aux données n'est pas autorisé depuis l'hôtel de ville.

### 3.2.5 Données d'utilisation des transports en commun

La société TEC possède des données d'utilisation des transports en commun sur l'ensemble de la commune de Charleroi à partir de 2016. En effet, depuis cette année-là, la TEC a mis en place un système permettant de compter le nombre d'usagers empruntant un transport TEC, que ce soit un bus ou un métro. De plus, le système permet à la TEC de connaître les types de carte d'embarquement. Il existe des cartes jetables, typiquement les billets valables pour un seul trajet, ainsi que des abonnements papiers ou des abonnements sans contact c'est-à-dire des abonnements utilisés sur une carte Mobib.

Les données fournies par la TEC donnent, pour chaque arrêt de bus ou de métro et pour une date et une heure précise au quart d'heure près, le nombre d'usagers qui ont pris le transport en commun pour un type de carte d'embarquement utilisé. Ces données sont regroupées sous forme de fichiers Excel, un par mois pour chaque année. La table 3.2 montre un exemple fictif d'une ligne du fichier Excel contenant les données du mois de janvier 2019. Cette table montre que le 29 janvier 2019 à 16h15, 17 voyageurs ont pris le bus à l'arrêt "Place de la Digue".

Vu que chaque ligne d'un fichier représente le nombre de voyageurs qui ont pris la TEC à un moment précis et pour un type de carte d'embarquement précis, le nombre total de lignes est très grand. Chaque fichier contient environ 90 000 lignes de données. Le total de lignes est donc estimé à environ 4 millions. Malheureusement, ces données ont été envoyées trop tard pour qu'une réflexion se fasse sur un besoin auquel une visualisation de ces données pourrait répondre pendant le stage.

TABLE 3.2 – Exemple fictif d'une ligne du fichier Excel pour le mois de janvier 2019. Cette ligne montre le nombre de voyageurs ayant pris un bus à l'arrêt "Place de la Digue" à 16h15 le 29 janvier.

Jour	Tranche quart d'heure	Libelle lieu	Nb Voyageurs	Libelle Type Support
29/01/19	16 :15	CHARLEROI Place de la Digue	17	Carte Sans Contact

### 3.2.6 Données sur les agents de la Ville

Le service de la Gestion des Ressources Humaines (GRH) de la Ville possède un certain nombre d'informations personnelles sur les quelques 3500 personnes travaillant pour la Ville,

appelées agents dans la suite de ce mémoire. Parmi ces agents, il y a, pour ne citer qu’eux, une majorité d’enseignants, des secrétaires, des agents travaillant en urbanisme ou encore dans l’état civil. Parmi ces données figurent les adresses des domiciles des agents.

Le Règlement Général sur la Protection des Données impose une gestion très particulière de ces données personnelles. Dès lors, les adresses des agents ne sont pas directement disponibles, c’est-à-dire que leur rue ainsi que leur localité sont correctes tandis que le numéro de maison n’est pas leur vrai numéro. Le faux numéro de maison donne une approximation du vrai numéro de maison car il est défini aléatoirement dans l’intervalle -20 à +20 par rapport au numéro réel. De plus, le nom de l’agent est aussi inconnu et est remplacé par un numéro de matricule qui permet de lier l’adresse à la bonne personne. Ce lien entre matricule et le nom de la personne concernée est uniquement connu du service GRH. Les 3500 adresses sont disponibles sous la forme d’un fichier CSV. Afin de compléter ces données, la Direction des Services Informatiques a fourni les adresses de l’ensemble des lieux de travail des agents.

La table 3.3 reprend la structure de ce fichier et donne un exemple fictif de données. La colonne **Service** reprend le service dans lequel l’agent travaille, la colonne **Cité** dit si l’agent travaillera dans la cité administrative.

TABLE 3.3 – Exemple montrant une adresse fictive d’un agent ainsi que diverses informations sur cet agent, comme le service dans lequel il travaille, si il travaillera dans la cité administrative et son matricule qui l’identifie.

Service	Cité	Rue	Numéro fictif	Code postal	Localité	Matricule
Systèmes d’informations	Oui	Rue des Bouvreuils	11	6534	Thuin	2746

### 3.2.7 Données KML provenant de cartes Google Maps

*Keyhole Markup Language*, KML dans la suite de ce mémoire, est un langage dérivé de XML dont l’objectif est la gestion de l’affichage de notations et de visualisations géographiques. Google MyMaps permet à ses utilisateurs de créer des cartes Google Maps personnalisées via l’insertion de symboles géographiques. Ces symboles peuvent être par exemple des points, des zones ou encore des lignes qui sont regroupés dans des calques en fonction de la classification que l’utilisateur veut effectuer sur ses données. Ces symboles peuvent être liés à des tableaux de données. Cette carte peut alors être partagée avec d’autres utilisateurs ou être exportée en fichier KML qui peut par la suite être utilisé dans une application tierce.

Cette exportation peut se faire sous plusieurs formes mais la plus intéressante est sans doute ce que Google appelle un fichier KML de lien réseau. Ceci permet d’utiliser la carte définie sur Google MyMaps dans une autre application tout en gardant les données à jour si l’utilisateur les modifie sur Google MyMaps.

Ce système proposé par Google MyMaps est très proche d’un logiciel SIG comme QGIS. Cependant il existe des différences entre les deux. En effet, Google MyMaps est très simple à utiliser et ne nécessite pas de logiciel SIG pour créer des cartes et les partager avec d’autres utilisateurs. En revanche, si un utilisateur veut réaliser des analyses poussées de données SIG, Google MyMaps peut devenir insuffisant. Dans ce cas, l’utilisateur peut privilégier des logiciels comme QGIS.

Plusieurs personnes travaillant en mobilité à la Ville utilisent régulièrement des cartes Google MyMaps afin de travailler sur différents dossiers nécessitant la définition de données géographiques sans pour autant avoir besoin d’utiliser QGIS. Par exemple, la figure 3.2 présente un calque de données représentant la position de garages pour vélos présents en ville.



FIGURE 3.2 – Exemple de calque créé sur Google MyMaps représentant un ensemble de garages pour vélos présents dans la Ville.

### 3.3 Inventaire des besoins et difficultés rencontrées

L’objectif de cette section est de lister l’ensemble des besoins découverts au début du stage afin de définir lequel de ces besoins servira de base à l’élaboration de la question de recherche. Les sections suivantes discutent chacune d’un besoin en particulier.

#### 3.3.1 Gestion des parkings

La circulation automobile est un problème récurrent en ville. Elle implique un besoin toujours plus grand de parkings. Or ce n’est pas possible d’en construire éternellement de nouveaux. De manière générale, les parkings sont divisés en trois catégories : les parkings publics détenus par la Ville, les parkings privés de moins de dix places et les parkings privés de plus de dix places comme par exemple les parkings des magasins. Les parkings privés de plus de dix places doivent payer une taxe annuelle à la Ville.

Une solution pourrait être d’optimiser l’utilisation de l’ensemble des parkings privés et publics présents en ville. Dans cette optique, la Ville a comme projet de réaliser un partenariat avec des sociétés privées possédant des parkings de plus de dix places. Ce partenariat serait le suivant. La Ville retire l’obligation de payer une taxe annuelle aux privés possédant des parkings en échange de quoi ces privés autorisent des covoitureurs à se parquer chez eux.

En 2018, un étudiant ingénieur de gestion a réalisé un mémoire portant sur la position et la priorisation de parkings de covoiturage à mettre en place par la Ville. Les résultats de ce mémoire révèlent des positions théoriques de parkings de covoiturage ainsi que le nombre de places à leur allouer. La Ville aimerait comparer les résultats du mémoire avec les données disponibles sur les parkings comme leur position et leur nombre de places. L’objectif est de définir quels parkings privés sont les mieux placés et les plus intéressants pour devenir des parkings de covoiturages.

Malheureusement, après quelques recherches, il s’avère que ce projet se rapproche plus d’un travail de gestion plutôt qu’un travail d’informaticien. En effet, il n’est pas utile de générer de nouvelles données ou une quelconque visualisation mais plutôt d’analyser les données existantes d’un point de vue de gestionnaire afin de définir des parkings de covoiturage. Ce besoin n’est donc pas choisi comme sujet de stage.

#### 3.3.2 Confrontation de FLOWCheck avec les données radars de la police

FLOWCheck permet à l’utilisateur de réaliser une analyse fine des routes empruntées par les voitures via notamment les segments de route. Cependant cette analyse de segments met uniquement en avant des tendances en pourcentage. La figure 3.1 présente un exemple de ce type d’analyse. Une couleur est attribuée à chaque segment de route en fonction du



pourcentage de voitures ayant emprunté ce segment. Plus ce pourcentage est grand et plus la couleur tend vers le rouge. En revanche, plus ce pourcentage est petit, plus la couleur tend vers le bleu et le blanc.

Un inconvénient de cette manière de présenter les choses est que le nombre réel de voitures est inconnu, ce qui pose problème aux experts utilisant cette plateforme car ils ne savent pas si les données sont représentatives ou pas. Par exemple, 80% à 90% des voitures passant par l'endroit choisi au préalable passent par un segment en particulier. Ce pourcentage n'a pas la même signification s'il représente 5 ou 500 voitures. Dans le premier cas, la route est peut-être une petite route de village et dans le second elle peut être un segment d'autoroute.

La Ville aimerait avoir une estimation du nombre réel de voitures qui empruntent ces segments de routes. Or la police possède des données recueillies par l'ensemble des radars positionnés en ville ainsi que par des installations de type DuoCount permettant de compter le nombre de voitures passant à l'endroit où l'installation est positionnée. Ces données sont présentées dans la section 3.2.4. Ce type de données permettrait de recueillir des informations précises sur la circulation et le nombre réel de voitures.

Une idée de solution pourrait être la définition d'une visualisation montrant le nombre de voitures circulant à différents endroits de Charleroi. Cette visualisation permettrait aux experts de confronter les données radars de la police avec les données de tendances de FLOWCheck.

Cependant la police dispose déjà d'un outil de visualisation de ses données radars. De plus, les données radars ne sont accessibles que par la police car ce sont des données sensibles. Il est seulement possible de recueillir certaines données sur une zone et une période particulière ainsi que les cartes statiques générées par l'outil de visualisation de la police. Cette piste de solution a donc été abandonnée.

### **3.3.3 Confrontation de FLOWCheck avec les choix proposés par les GPS**

Grâce à FLOWCheck, la Ville peut avoir des informations sur les trajets empruntés par les conducteurs via l'analyse de segments de route. Cependant, elle ne sait pas exactement pourquoi les conducteurs prennent ces chemins et pas d'autres. Cela peut poser problèmes dans certaines circonstances. Par exemple, des petites rues sont de plus en plus fréquentées par les voitures à certains endroits de l'entité de Charleroi car les GPS comme Waze ou Google Maps proposent des trajets en fonction du trafic en temps réel ce qui a pour conséquence, lorsque le trajet traditionnel est embouteillé, de faire passer les voitures par des rues qui ne sont pas prévues pour recevoir autant de trafic. La mobilité de la Ville est tributaire des trajets proposés par les différents GPS.

Dans ce contexte, la Ville aimerait confronter les données proposées par FLOWCheck et les données proposées par les GPS afin d'avoir une vision la plus complète possible de la mobilité routière au sein de la commune. Malheureusement FLOWCheck ne fournit pas d'API pour accéder aux données générées lors des requêtes des utilisateurs. Les données sont bien disponibles mais uniquement manuellement. Si un expert veut accéder à ces données, il doit les télécharger à partir des résultats générés par FLOWCheck. Cela implique que l'expert doit savoir exactement quelles données il veut avoir et donc quelles données sont importantes. Cela n'a pas beaucoup de sens dans un contexte comme celui-ci où l'expert ne sait pas ce qu'il cherche. Pour ces raisons, le projet tel quel n'a pas été retenu.

### **3.3.4 Affinage des résultats de FLOWCheck**

Un inconvénient de l'analyse de segments de route de FLOWCheck qui n'a pas encore été discuté est sa visualisation confuse. En effet dans le contexte du trafic celle-ci ressemble très fort à une *flow map* représentant la densité de circulation des routes à un instant donné. Les

experts doivent fournir un effort cognitif important pour empêcher leur esprit d’interpréter la visualisation de cette manière.

De plus il est difficile d’analyser l’évolution des données à travers les années. Pour rappel la Ville a accès à des données de janvier 2017 jusqu’en décembre 2021. Cependant, elle ne peut comparer facilement que les données issues d’une même année car FLOWCheck a fourni des identifiants de connexion différents pour chaque année, ce qui rend très compliqué la comparaison entre des mêmes segments à plusieurs années d’intervalle.

Une plateforme pourrait être développée afin de regrouper les segments jugés importants et compléter l’analyse des segments de route de FLOWCheck via une visualisation de l’évolution de ces segments à travers le temps. Cette visualisation devrait être la plus claire possible afin de réduire au maximum l’effort cognitif fourni par les experts et qu’ils puissent se concentrer sur l’analyse des données sans perdre de temps. Malheureusement il y a un problème rencontré, le même que dans la section 3.3.3. FLOWCheck ne fournit pas d’API pour accéder à ses données. Il est dès lors plus compliqué de récupérer efficacement les données de circulation. Pour se faire, il est nécessaire de savoir distinguer les données importantes des données essentielles. Or les experts ne savent décider quelles données sont essentielles et lesquelles ne le sont pas qu’après avoir analysé les données. Pour ces raisons, cette idée de projet n’a pas été retenue.

### 3.3.5 Simplification de l’utilisation de FLOWCheck

FLOWCheck est un outil intéressant qui peut fournir un très grand nombre d’analyses de circulation en fonction des demandes des experts. Cependant comme vu dans les sections précédentes, il ne peut être utilisé au maximum de son potentiel que si les experts savent ce qu’ils cherchent c’est-à-dire quelle route ou segment de route analyser ainsi que la période et les tranches horaires intéressantes.

La Ville aimerait trouver un moyen d’aider les experts à recueillir les informations préalables dont ils ont besoin pour ensuite utiliser FLOWCheck. Une plateforme pourrait être développée pour répondre à ce besoin en visualisant des données de circulation en temps réel ainsi que des données sur les transports en commun comme la ponctualité des transports ainsi que le nombre d’utilisateurs en fonction de l’heure et des arrêts. Cette visualisation offrirait un nouveau point de vue sur la circulation à Charleroi de par les données différentes qu’elle proposerait. En effet, là où FLOWCheck donne des données historiques sur la circulation automobile, cette plateforme proposerait une visualisation de données en temps réel sur le trafic ainsi que des données historiques sur l’utilisation des transports en commun qui participent aussi à la circulation en ville. Ce point de vue nouveau pourrait aider les experts à définir des zones d’intérêts à des moments particuliers pour les analyser plus en détail dans FLOWCheck par la suite.

Les données de ponctualité des trains sont fournies par Infrabel via une API. Cependant il n’existe pas encore d’équivalent chez TEC. Or les bus et le métro sont les transports en commun majoritaires à Charleroi. Les données discutées dans la section 3.2.5 ont été reçues très tard dans le stage et n’étaient donc pas encore disponibles quand la question de recherche a été définie. Pour toutes ces raisons, ce projet n’a pas été retenu.

### 3.3.6 Création d’une carte interactive de mobilité

La ville d’Ottignies-Louvain-la-Neuve possède une carte interactive destinée aux citoyens de la commune et leur donne facilement accès à des informations sur la mobilité de la ville comme par exemple les lignes de bus avec leurs horaires, leurs emplacements ou encore les aménagements cyclables. La figure 3.3 montre un exemple tiré de cette carte qui visualise la ligne de bus 31 ainsi que les aménagements cyclables de la ville.

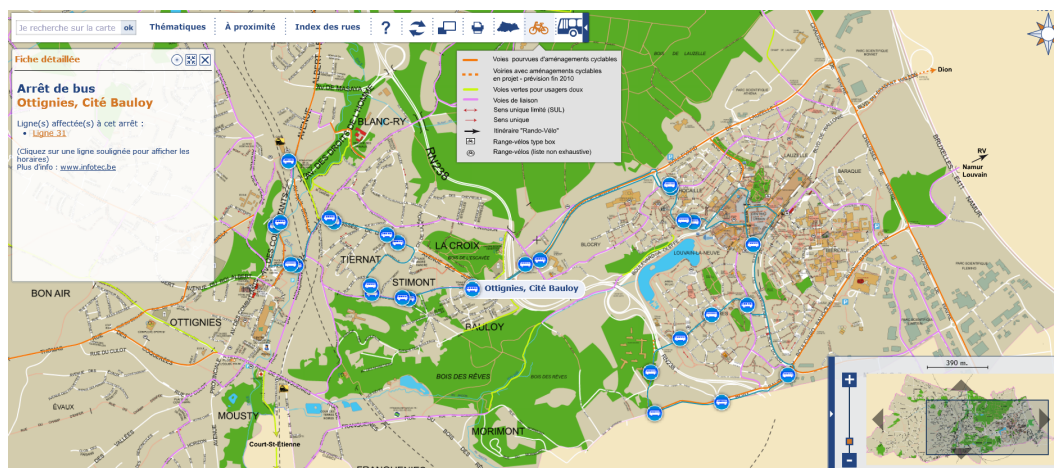


FIGURE 3.3 – Carte interactive de mobilité de la ville d’Ottignies-Louvain-la-Neuve montrant ici la ligne de bus 31 ainsi que les aménagements cyclables de l’ensemble de la commune (carte disponible à l’adresse suivante : <http://map.olln.be/>).

La ville de Charleroi aimerait se doter d’un outil similaire qui permettrait de centraliser l’ensemble des données de mobilité dont la Ville dispose et qui serait mis à disposition des carolorégiens ainsi que de différents services de la Ville. L’outil serait donc divisé en deux parties. La partie destinée aux citoyens serait similaire à la carte interactive présentée ci-dessus et leur présenterait donc les données de la Ville sans qu’ils ne puissent les modifier. La partie destinée aux services en lien avec la mobilité permettrait à ceux-ci de modifier les données ainsi qu’une meilleure centralisation de celles-ci, ce qui simplifierait le travail effectué en mobilité par ces services.

Cependant plusieurs problèmes se posent. Tout d’abord, la Ville ne possède pas de données dynamiques intéressantes à intégrer dans une carte interactive qui pourraient être utilisées par après par le cabinet de mobilité. Dans ces conditions, l’utilisation de QGIS pour analyser ces données statiques et les visualiser sur une carte est suffisante.

Ensuite, la Ville va bientôt se munir d’une solution ArcGIS qui permet de modifier des données tout en les centralisant. Donc l’idée de réaliser une plateforme à destination des services de la Ville afin de centraliser les données n’est plus pertinente.

La carte interactive à destination des citoyens reste d’actualité. Cependant, vu le peu de données dont la Ville dispose, ce projet n’a pas beaucoup d’intérêt pour le moment. Si dans les mois ou années qui viennent la Ville se munit de nouvelles données alors il sera peut-être intéressant de se pencher sur la problématique de la carte interactive.

### 3.3.7 Gestion des déplacements domicile-travail des agents

Une partie des agents de la Ville est répartie dans différents services situés à divers endroits de l’entité administrative de Charleroi. La Ville aimerait avoir une idée des déplacements domicile-travail de ses agents afin de remplir deux objectifs sur le long terme.

Le premier objectif est de remplir plus facilement l’Obligation Fédérale de Diagnostic, un document envoyé à chaque entreprise privée ou publique comptant plus de 100 travailleurs. Ce document, à remplir tous les trois ans, permet au Service Public Fédéral Mobilité et Transports d’avoir notamment une idée des déplacements domicile-travail des travailleurs en Belgique afin d’adapter les transports<sup>2</sup>. Or pour répondre au questionnaire fédéral, la Ville doit connaître les moyens de déplacement de ses agents.

Le deuxième objectif est de proposer des moyens de transport plus durable à ses agents en privilégiant les transports en commun, le vélo et la marche à l’utilisation d’une voiture

2. Voir le lien suivant pour plus d’informations à ce sujet : <https://www.mobilite-entreprise.be/index.php/mobility-management/lobligation-federale-de-diagnostic/>

personnelle dans le cadre des trajets domicile-travail de chaque agent.

Dans ce contexte, la Ville a pour projet de rassembler plusieurs de ses services en un seul endroit situé au centre-ville : la cité administrative. À terme, la cité administrative regroupera environ 650 des 3500 agents. Dans un premier temps, la Ville aimerait réaliser ces deux objectifs dans le cadre de la cité administrative et puis dans un second temps élargir à l'ensemble des 3500 agents.

Plusieurs problèmes se posent dans cette définition de besoin. Le plus important est le manque de données concernant les habitudes de transport des agents. En effet, la Ville ne dispose que des adresses des domiciles des agents mais n'a aucune donnée concernant les trajets effectués ainsi que le mode de transport utilisé entre le domicile et l'endroit de travail. Un deuxième problème qui mérite d'être soulevé est la protection de la vie privée. En effet, le RGPD empêche la Ville de fournir des informations personnelles sur ses agents à n'importe qui. Or les adresses des domiciles révèlent des informations personnelles sur les agents.

Malgré ces problèmes, la problématique de la gestion des déplacements domicile-travail des agents a été choisie pour le stage. Les deux objectifs poursuivis par la ville ont permis de définir la question de recherche qui est, pour rappel, la suivante.

Quelles techniques de visualisation de l'information peuvent aider la ville de Charleroi à connaître les déplacements domicile-travail de ses agents et à leur proposer un mode de transport adapté à leurs déplacements et le plus durable possible ?

### 3.4 Développement d'une solution concernant les déplacements domicile-travail des agents

La question de recherche a été définie pour répondre à la problématique soulevée dans la section 3.3.7. Cette section-ci regroupe l'explication du développement d'une solution de visualisation permettant de répondre à cette question de recherche. La section 3.4.1 décrit l'analyse de la problématique et sa division en plusieurs sous-problèmes permettant de définir l'architecture de la solution à la section 3.4.3. L'ensemble des données provenant d'API est discuté dans la section 3.4.2. La section 3.4.4 discute du *preprocessing* réalisé sur les données utilisées. La section 3.4.5 présente le développement des différentes visualisations répondant à la question de recherche. L'utilisation de la solution développée implique un certain coût qui est discuté dans la section 3.4.6. Enfin, la section 3.4.7 présente l'évaluation de la solution.

#### 3.4.1 Analyse de la problématique

La question de recherche se base sur les besoins évoqués dans la section 3.3.7. Les deux objectifs définis dans cette section nécessitent de diviser le projet de développement de visualisation en deux.

Une première partie est destinée aux experts de la Ville qui travaillent et prennent des décisions de mobilité. Elle a pour objectif de donner des informations sur les déplacements des quelques 650 agents qui travailleront à la cité administrative entre leur domicile et leur futur lieu de travail en fonction de leur moyen de transport. Pour ce faire, il est nécessaire d'avoir une vision large des flux de déplacements des agents ainsi qu'une vision des trajets potentiels proposés par Google Maps.

La vision globale donne une vue générale de la répartition des agents sur le territoire ainsi que de leurs flux OD. La vision des trajets potentiels quant à elle propose une visualisation des déplacements des agents en fonction de leurs modes de transport. Elle permet de visualiser d'éventuels *patterns* qui se dégageraient des données afin d'adapter l'offre de la Ville comme par exemple positionner des parkings de covoiturage en fonction de ces *patterns*.

La deuxième partie est destinée à l'ensemble des agents de la Ville et a pour objectif de montrer à chacun d'eux une comparaison entre différents moyens de transport pour réaliser le trajet de son domicile vers son lieu de travail. L'idée sous-jacente de cette partie est de pousser les agents à modifier leur mode de transport pour une solution plus durable.

### 3.4.2 Données en temps réel provenant d'API

Les données fournies par le service GRH et discutées dans la section 3.2.6 ne donnent que des informations sur la position des domiciles et rien d'autre. Elles ne sont donc pas suffisantes pour répondre à la problématique de la Ville. Pour ce faire, il est nécessaire d'avoir accès à d'autres types de données. Ces données sont décrites dans cette section.

Il s'agit de données de mobilité en temps réel fournies par des API provenant de différentes sources comme Google Maps, le GPS TomTom et Mapbox. Bien que Google Maps et TomTom soient connus et n'aient pas besoin d'être présentés, le dernier en revanche n'est pas connu du grand public. Mapbox est une entreprise fournissant des outils de cartographie en ligne et *open source*. Les données provenant de ces trois entreprises peuvent être très intéressantes car elles donnent des informations précieuses. Ces informations portent, par exemple, sur la localisation d'événements sur une carte, via une API de géolocalisation, ou encore sur les chemins empruntés par les usagers sur la route en fonction de leur mode de transport, via une API de directions.

**API de géolocalisation** Cette API, que ce soit pour Mapbox, TomTom ou Google Maps, permet de récupérer les coordonnées d'une adresse donnée, ce qui donne la possibilité de les placer sur une carte. Les API de Mapbox, TomTom et Google Maps ont été testées afin de sélectionner la plus efficace. L'efficacité d'une de ces API est définie en fonction de sa capacité à récupérer les bonnes coordonnées d'une adresse.

Plusieurs raisons peuvent être à l'origine de mauvaises coordonnées retournées par l'API. Celle-ci peut tout d'abord ne pas connaître la rue. Dans ce cas, elle peut retourner les coordonnées du centre de la localité de l'adresse demandée. Elle peut aussi connaître la rue mais pas le numéro, auquel cas elle donnera les coordonnées du centre de la rue. Dans ces deux cas, les coordonnées retournées sont considérées comme mauvaises car elle ne sont pas assez précises. Il est aussi possible que l'API retourne des coordonnées qui correspondent à la rue demandée mais pour une autre localité.

Voici les résultats des appels à ces API pour les 658 adresses des agents travaillant à la cité administrative.

1. Mapbox : 602 adresses ont été bien positionnées. Cela donne un taux de réussite de 91,5%.
2. TomTom : 523 adresses ont été bien positionnées. Cela donne un taux de réussite de 79,5%.
3. Google Maps : 651 adresses ont été bien positionnées. Cela donne un taux de réussite de 98,9%.

L'API de Google Maps, *Geocoding API*, est la plus performante des trois. C'est donc cette API qui sera utilisée.

**API de directions** Mapbox fournit également une API de directions. Au delà d'un certain seuil d'utilisation de l'API, que ce soit pour Mapbox, TomTom ou Google, il faut payer plus ou moins en fonction de l'API sur laquelle la requête se fait et du nombre de requêtes. Afin de choisir l'API la plus intéressante parmi les trois à disposition, il faut pouvoir les comparer en terme de fonctionnalités et de prix. Après comparaison entre les trois API via la table 3.4, l'API Google Maps, *Directions API*, est choisie car elle fournit plus de données que les autres.

TABLE 3.4 – Comparaison des API de directions de Mapbox, TomTom et Google Maps

	Mapbox	TomTom	Google Maps
Avantages	- Open Source	- Précision des résultats	<ul style="list-style-type: none"> <li>- Facilité d'utilisation</li> <li>- Précision des résultats</li> <li>- Multi modalité (voiture, bus, tram, métro, train, vélo, marche)</li> <li>- Facilité d'intégration des résultats sur une carte</li> </ul>
Inconvénients	- Aucun transport en commun	<ul style="list-style-type: none"> <li>- Difficile à utiliser</li> <li>- Difficile à intégrer sur une carte</li> <li>- Pas de train, tram, metro</li> <li>- Pas de multi modalité</li> </ul>	- Plus cher
Nombre de requêtes gratuites par mois	100 000	Environ 75 000	Environ 40 000
Prix pour 1000 requêtes	2\$ (de 100 000 à 500 000 requêtes)	0.5\$	5\$

**API de dénivelé** Google Maps fournit une API permettant de calculer le dénivelé d'un trajet entre un point A et B nommée *Elevation API*. Etant donné que les API de géolocalisation et de directions utilisées sont celles proposées par Google Maps, l'API de dénivelé utilisée pendant le développement est celle de Google Maps ce qui permet une simplification de la mise en production du produit final.

**API Google Charts** Afin d'utiliser les données récupérées à partir de l'API *elevation* de Google Maps, il est nécessaire de pouvoir générer des graphiques. Il existe plusieurs librairies permettant de le faire comme *D3.js* ou *C3.js* qui est un *wrapper* de *D3.js*. Il existe également des API comme *Charts API* qui est fournie par Google. C'est cette dernière qui sera utilisée pour la génération de graphiques. En effet, cela permet de ne pas devoir intégrer une nouvelle librairie dans le projet. Plus il y a de librairies utilisées, plus un projet peut être lourd et plus il y a de chances que les librairies ne fonctionnent pas entre elles, ce qui amènerait des *bugs*.

**API Places** Google Maps fournit une API, *Places API*, qui retourne une liste d'adresses Google Maps en fonction d'une adresse spécifiée dans la requête. Comme pour l'API de dénivelé, l'API *Places* de Google Maps sera utilisée lors du développement de la solution car cela simplifiera la mise en production de la plateforme.

**API Maps** Mapbox, TomTom et Google Maps proposent tous une API de cartes géographiques. Comme pour l'API *Places* ainsi que l'API *Elevation*, l'API *Maps* de Google sera utilisée afin de faciliter le déploiement de la solution pour les clients.

### 3.4.3 Architecture

**Choix de la technologie** Afin de répondre à la première problématique soulevée dans la section 3.4.1, plusieurs pistes ont été empruntées.

La première était d’implémenter une visualisation qui s’intégrerait dans le logiciel QGIS afin de permettre aux personnes travaillant avec ce logiciel de les utiliser directement. Cependant cette piste a vite été abandonnée car après quelques recherches, aucune information sur une quelconque possibilité d’implémenter de nouvelles visualisations n’a été trouvée.

La deuxième piste était une implémentation sous forme d’une plateforme web. En effet, de nombreuses bibliothèques Javascript permettent d’implémenter des visualisations de tous types tout en ajoutant de l’interaction entre l’utilisateur et la visualisation. De plus, il existe aussi des bibliothèques permettant de réaliser une interface utilisateur ergonomique. Le développement de la solution proposée est donc réalisé sous forme de deux sites web répondant aux deux problématiques définies dans la section 3.4.1. Les deux plateformes sont développées en Javascript.

**Choix des bibliothèques** Concernant l’implémentation des visualisations, différentes solutions étaient possibles. Dans la plateforme destinée à la Ville, des *flow maps* ont été implémentées et l’une d’elles, la visualisation présentée à la figure 3.8, a nécessité l’utilisation d’une bibliothèque.

Deux bibliothèques ont été envisagées : `D3.js` et `flowmap.gl`. Elles permettent toutes deux d’implémenter des visualisations cartographiques via des cartes Mapbox.

`D3.js` est une bibliothèque permettant d’implémenter de nombreuses visualisations et a pour avantage un très grand contrôle du résultat visuel de celles-ci (Bostock et al., 2011). Cependant, la conséquence est que l’implémentation de visualisations est plus compliquée qu’avec d’autres bibliothèques.

De son côté, `flowmap.gl` est, comme son nom l’indique, une bibliothèque spécialisée dans la visualisation de *flow maps*. Elle est moins puissante que `D3.js` car elle ne permet pas de réaliser autant de visualisations différentes que celui-ci et ne donne pas de contrôle aussi précis sur les visualisations. Cependant il n’est pas nécessaire d’avoir un contrôle important sur cette visualisation. Cette dernière a donc été implémentée avec `flowmap.gl`.

Dans la partie destinée à l’ensemble des agents de la Ville, l’ensemble des données utilisées provient des API de Google Maps. Dès lors, l’API Google Maps Javascript a été utilisée car elle intègre les API utilisées et discutées dans la section 3.4.2.

Concernant l’interface graphique, l’utilisation d’une bibliothèque permet de simplifier grandement le développement tout en proposant une interface professionnelle. Il existe plusieurs bibliothèques comme `VueJS`, `ReactJS` et `AngularJS`. Chacune d’elles a ses avantages et ses inconvénients. Cependant la bibliothèque `flowmap.gl` a été implémentée pour s’intégrer à `ReactJS`. Cette dernière a donc été choisie pour l’implémentation de l’interface graphique et plus particulièrement `material-ui` qui fournit des composants graphiques prêts à l’emploi et qui facilite le développement de l’interface.

**Structure des plateformes web** La figure 3.4 montre la structure des fichiers de la plateforme web à destination des experts. Ce mémoire n’a pas pour objet le fonctionnement de `ReactJS` qui n’est donc pas expliqué en profondeur ici. Ce qu’il est important de retenir est que React fournit un fichier `App.js`. Le rendu de ce *frontend* sera effectué à partir de ce fichier. Celui-ci ouvre deux fichiers, un pour chaque visualisation. `FlowMovement.js` s’occupe de la visualisation du mouvement global des agents ainsi que de la visualisation de la position de leur domicile. Il utilise comme sources de données deux fichiers JSON : `network.json`, qui contient des *clusters*, ainsi que `agents.json` qui contient des données sur les 650 agents. Ces deux fichiers JSON sont définis dans la section 3.4.4. De son côté, `PotentialTrajectories.js` s’occupe de la visualisation des trajets potentiels proposés par Google Maps. Il utilise le fichier `WeightedTrajectories.json` comme source de données. Ce fichier JSON reprend des trajets et est défini dans la section 3.4.4. Enfin, les deux fichiers `FlowMovement.js` et `PotentialTrajectories.js` font appel à divers fichiers de style CSS.

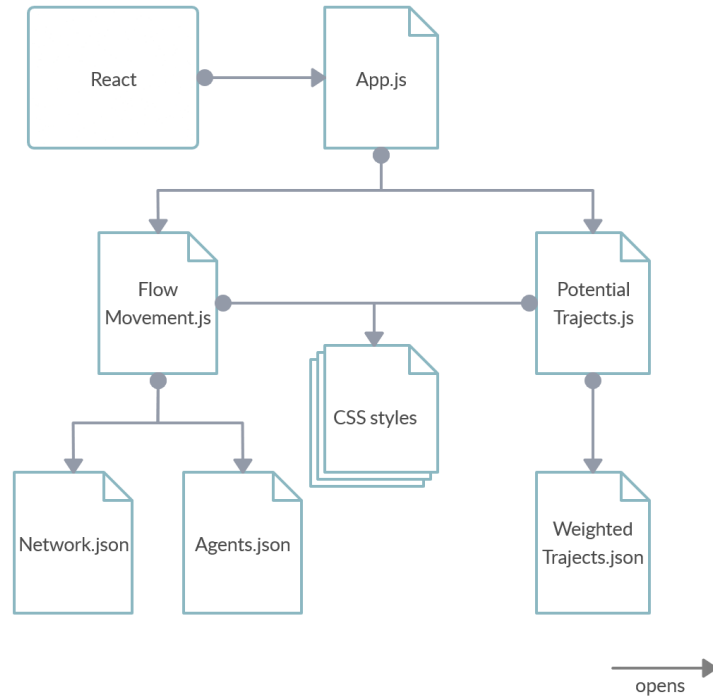


FIGURE 3.4 – Structure des fichiers de la plateforme web à destination des experts montrant la hiérarchie des fichiers.

La figure 3.5 montre la structure des fichiers de la plateforme web à destination des agents de la Ville. La structure générale est similaire à celle de la plateforme à destination de la Ville. Le fichier **TrajectAgents.js** reprend l'ensemble de la visualisation. Il effectue des appels aux différentes API de Google. Certains résultats des appels aux API sont directement utilisés dans le fichier. D'autres sont envoyés à **SearchBox.js** qui s'occupe de la gestion d'un formulaire de recherche mis à disposition des utilisateurs. **SearchBox.js** utilise un autre fichier JSON nommé **CoordBuildings.json** qui contient les coordonnées des différents lieux de travail. Ce fichier est défini dans la section 3.4.4.

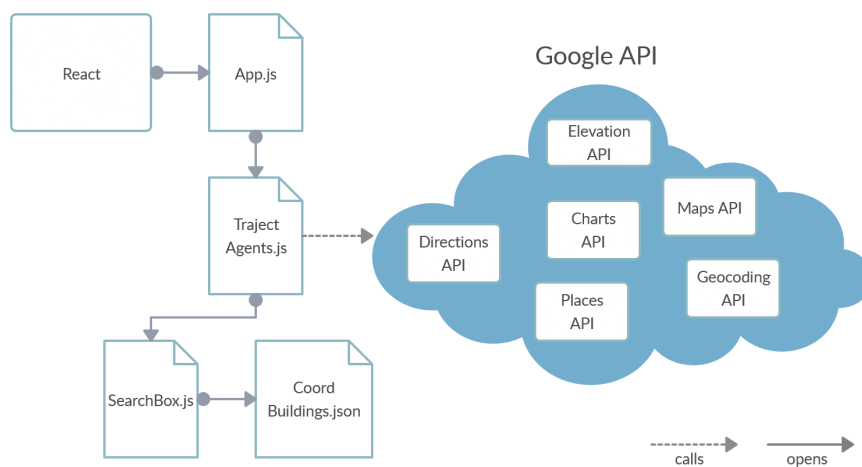


FIGURE 3.5 – Structure de la plateforme web à destination des agents montrant la hiérarchie des fichiers ainsi que la manière dont les API sont appelées.



### 3.4.4 *Preprocessing* des données

Les deux plateformes web se basent sur diverses données qui sont traitées au préalable en Python. L'explication de ces traitements est décrite ci-dessous.

**Plateforme à destination des experts** Le *preprocessing* des données utilisées pour cette plateforme est divisé en trois parties. La première discute des adresses qui seront utilisées dans les parties suivantes. La deuxième discute des données de la visualisation des flux globaux et la troisième discute des données de la visualisation des trajets potentiels proposés par Google Maps.

***Preprocessing* des adresses** Les données de base du site web à destination de la Ville sont les adresses postales des 650 agents décrites dans la section 3.2.6. Un premier traitement à effectuer est de transformer ces données textuelles en coordonnées géographiques grâce à l'API *Geocoding* de Google Maps décrite dans la section 3.4.2. Les adresses postales des agents sont donc transformées en données spatiales. Ces données sont alors insérées dans un fichier JSON appelé `Agents.json` qui est utilisé par la suite dans la visualisation des flux à destination de la Ville.

***Preprocessing* de la vision globale des flux** Afin de développer une *flow map* avec `flowmap.gl` telle que discutée dans la section 3.4.5, des zones agrégeant les positions des domiciles doivent être spécifiées et des flux doivent être définis entre ces zones. Cette partie du *preprocessing* s'intéresse donc d'abord à la création des zones et puis à la définition des flux.

Les zones Proximus discutées dans la section 3.2.3 sont très petites et permettent d'agréger moins d'une dizaine d'adresses par zone. Elles ne sont donc pas pertinentes. En effet, comme discuté dans la section 2.5, une *flow map* n'est intéressante que si elle permet une analyse macro des données, or une telle analyse n'est possible que lorsque les données sont agrégées dans de plus grandes zones que celles proposées par Proximus. Il est donc nécessaire de générer de plus grands *clusters* rassemblant les données.

Il est à noter que dans un souci de lisibilité, pour les deux techniques de *clustering* discutées ci-dessous, les domiciles situés à moins de deux kilomètres de la cité administrative, c'est-à-dire 68 domiciles, ont été retirés des données. Cette lisibilité est discutée en détail dans la section 3.4.5. L'hypothèse faite est donc que ces données retirées forment un *cluster* représentant l'ensemble des domiciles se trouvant au centre-ville. Sur les 650 domiciles, seuls 582 sont donc considérés dans les algorithmes de *clustering* suivants.

Différentes techniques de *partitional clustering* peuvent être utilisées. Les deux principales sont le *density-based spatial clustering of applications with noise* (DBSCAN) et *k-means*. DBSCAN utilise deux paramètres : une distance  $\epsilon$  et le nombre minimal de points qui doivent se trouver dans un rayon  $\epsilon$  pour former un *cluster*. *k-means* utilise un seul paramètre,  $k$ , qui définit le nombre de *clusters* qui doivent être générés. Ces deux algorithmes ont été testés afin de choisir le plus efficace et pertinent.

En théorie, DBSCAN est tout indiqué pour générer des *clusters* en fonction de la densité des domiciles des agents. Cependant la figure 3.6 montre qu'en pratique, un seul *cluster* est généré à partir des 582 points. Le test a été reproduit plusieurs fois en changeant la valeur des deux paramètres. L'algorithme ne trouve à chaque fois qu'un seul *cluster* avec plus ou moins d'*outliers* en fonction de la valeur des paramètres.

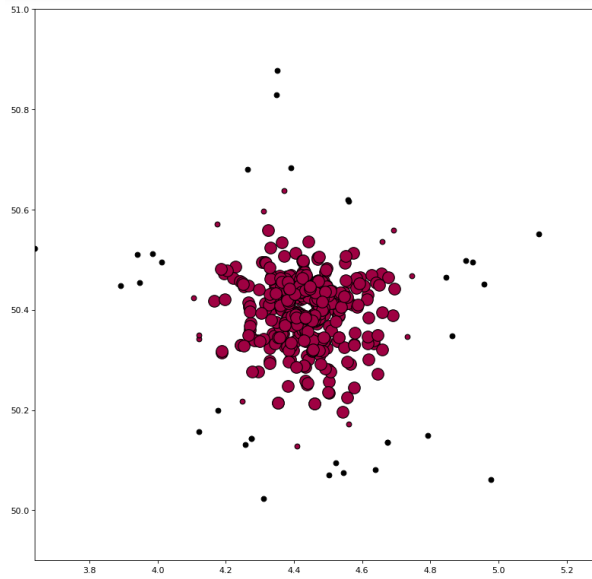


FIGURE 3.6 – Résultat d'un *partitional clustering* des coordonnées des domiciles des 650 agents via la technique DBSCAN ayant comme paramètres  $\text{distance}=0.1$  et  $\text{min-samples}=10$ . Les points rouges forment un *cluster* et les points noirs sont considérés comme *outliers*. Les axes représentent les coordonnées géographiques des points.

L'algorithme *k-means* quant à lui donne des résultats plus intéressants. Après plusieurs essais, la valeur de  $k$  la plus pertinente est 20. La figure 3.7 montre le résultat du partitionnement des coordonnées en 20 *clusters*. Chaque couleur représente un *cluster* et chaque point gris représente le centroïde d'un *cluster*. Cet algorithme a donc été choisi pour générer les *clusters*. Ceux-ci sont insérés dans un fichier JSON appelé `Network.json` qui est par après utilisé dans la *flow map* décrite dans la section 3.4.5.

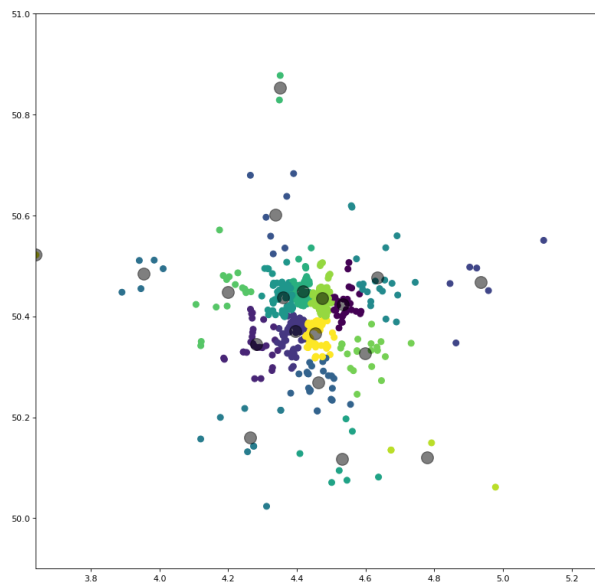


FIGURE 3.7 – Résultat d'un *partitional clustering* des coordonnées des domiciles des 650 agents via la technique *k-means* pour  $k=20$ . Chaque couleur représente un des 20 *clusters* et chaque point gris représente le centroïde d'un *cluster*. Les axes du graphe représentent les coordonnées géographiques des points.

Après avoir défini les zones, les flux entre ces zones peuvent être générés. Un script Javascript va définir ces flux en fonction du nombre d'adresses présentes dans chaque zone. Ces flux ont une zone d'origine, une zone de destination et le nombre d'adresses présentes dans la zone d'origine.

**Preprocessing de la vision des trajets potentiels** Afin d'implémenter la visualisation des trajets empruntés par les agents de la cité administrative, il est nécessaire de récupérer les trajets proposés par l'API Directions de Google Maps. Ces données sont définies comme des trajectoires partant d'un point de départ vers un point d'arrivée et sont associées à un mode de transport parmi les suivants : voiture, vélo, marche ou transports en commun. Le moyen de transport associé à un trajet est défini en fonction de la distance à vol d'oiseau entre le point de départ de ce trajet et la cité administrative. Les domiciles à moins de 2 km de la cité administrative auront des trajets à pied. Les domiciles entre 2 km et 5 km feront leur trajet en vélo et enfin les domiciles à plus de 5 km se verront attribuer la voiture ou les transports en commun comme moyen de transport.

Chaque trajet est divisé en plusieurs segments de route qui mis bout à bout forment le trajet en entier. Les trajets sont donc des données *nominal* et ne donnent que des informations sur quelles routes sont empruntées ou non et par quel moyen de transport. La question de recherche nécessite de recueillir des informations sur les tendances de déplacement des agents en fonction de leur moyen de transport. Pour pouvoir trouver ces informations, il est nécessaire de changer l'échelle de mesure des données. C'est l'objet de la deuxième partie de ce *preprocessing*.

Pour chaque moyen de transport, tous les trajets générés par Google Maps ont été agrégés en un seul grand réseau. Chaque segment de route de ce réseau a un poids défini en fonction du nombre de fois que ce segment est emprunté par un trajet. Si un segment A est emprunté par un seul trajet, il aura par exemple un poids de 0,01 tandis qu'un segment B emprunté par 20 trajets aura par exemple un poids de 0,2. Ces poids permettent une comparaison plus poussée des données en disant par exemple que le segment B est 20 fois plus emprunté que le segment A, par rapport à une échelle de mesure *nominal* capable d'indiquer uniquement si le segment est emprunté ou non. Ces données sont donc à présent des données de type *ratio* et permettent une définition de visualisations mettant en avant d'éventuels *patterns*. Elles sont insérées dans un fichier JSON appelé `WeightedTrajects.json`.

**Plateforme à destination des agents** La plupart des données utilisées dans cette plateforme viennent des API de Google qui sont déjà traitées et prêtes à l'utilisation. Les seules données qui doivent passer par une étape de *preprocessing* sont les adresses des différents lieux de travail. Pour pouvoir les utiliser, il est nécessaire de récupérer leurs coordonnées, ce qui est fait via l'API *Geocoding* de Google Maps. Les données textuelles de base sont donc transformées en données spatiales. Celles-ci sont alors insérées dans un fichier JSON appelé `CoordBuildings.json`.

### 3.4.5 Développement des visualisations

Après avoir identifié et transformé les données via du *preprocessing*, les visualisations ont été développées pour répondre à la question de recherche. Comme décrit dans la section 3.4.3, celles-ci sont divisées en deux : les visualisations à destination des experts en mobilité de la Ville et les visualisations à destination de l'ensemble des agents.

**Visualisations à destination des experts** Les visualisations destinées aux experts sont divisées en deux parties : une partie globale composée d'une visualisation des flux OD des agents et d'une visualisation des positions des domiciles ainsi qu'une partie composée d'une visualisation des trajets des agents. Vu l'absence de données temporelles, les visualisations à destination des experts sont basées sur des techniques de visualisation de l'espace.

Le développement de ces visualisations à destination des experts a été pensé pour suivre le mantra "*Overview first, zoom and filter, details on demand*" défini par Shneiderman. Ce mantra discute des besoins d'une vision globale et d'une vision précise d'une visualisation (Shneiderman, 1996).

Dans ce cadre, *overview first* signifie que l'utilisateur doit avoir une vision claire et entière des données. Cette vision peut être vue comme une sorte de résumé. Elle est ici représentée par la visualisation du mouvement global des agents ainsi que la visualisation de la position de leur domicile.

*Zoom and filter* fait référence aux parties de la visualisation qui donnent des informations précises sur un sous-ensemble des données. Ceci a été traduit par le *filtering* ajouté dans chacune des visualisations. Cette technique d'interaction permet aux experts de ne visualiser que les flux les plus importants ainsi que les trajets les plus importants en fonction du moyen de transport.

Enfin, la partie *details on demand* de la visualisation donne toute l'information d'un objet précis. Cette partie n'est pas présente dans les visualisations à destination des experts. En effet, le développement des visualisations à destination des experts a été interrompu au profit de la visualisation à destination des agents de la ville. Ceci est discuté plus en détail dans la section 3.4.7.

**Visualisation du mouvement global des agents** La visualisation du mouvement global propose une vision des mouvements domicile-travail des agents et fait donc partie de la famille des techniques *line-based*. En effet, elle est basée sur des flux OD qui peuvent être utilisés dans des techniques de cette famille. Une *flow map* représentant les déplacements des agents entre leur domicile et la cité administrative a été développée et est présentée à la figure 3.8.

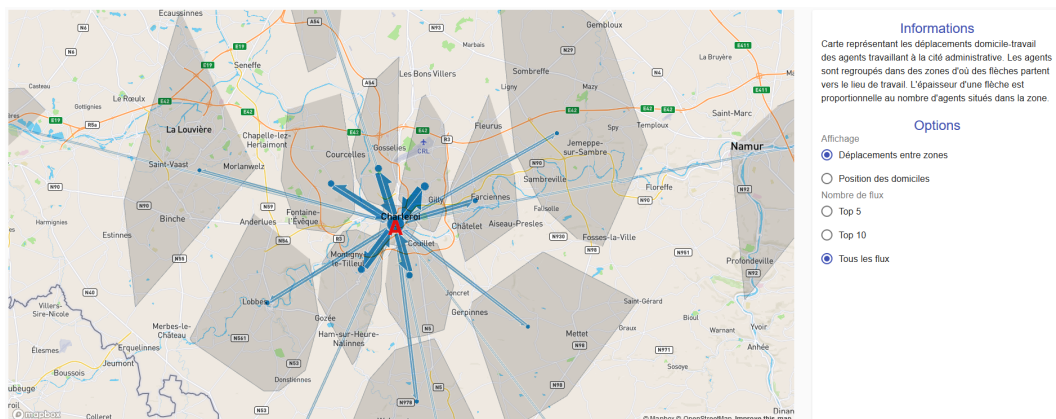


FIGURE 3.8 – Visualisation des flux entre chaque *cluster* et la cité administrative (A). Un polygone représente un *cluster* et les flux sont représentés par les flèches dont l'épaisseur est proportionnelle au nombre de domiciles présents dans le *cluster* correspondant.

Cette technique de visualisation a été choisie car elle est simple à comprendre et représente les informations directement sur la carte, contrairement à une *OD matrix* ou une *OD map* pour lesquels l'effort cognitif nécessaire pour lier les données à la carte est beaucoup plus grand. De plus, les données disponibles sont typiquement des données OD. Or une *flow map* permet une analyse macro de flux OD, ce qui est intéressant pour la Ville car cela peut lui révéler des tendances de déplacement domicile-travail.

Pour ce faire, cette *flow map* se base sur les *clusters* définis dans la section 3.4.4. Les déplacements entre chaque *cluster* et la cité administrative sont représentés par des flèches dont l'épaisseur est proportionnelle au nombre de domiciles présents dans le *cluster* correspondant. Avant de développer cette *flow map*, des *mockups* papier ont été réalisés afin d'avoir des retours de la part du client. La figure 3.9 montre un de ces *mockups*.

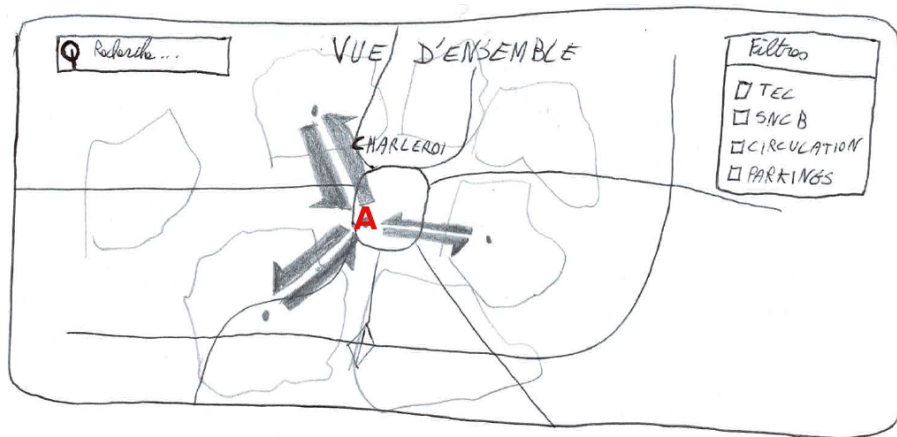


FIGURE 3.9 – Mockup de la visualisation des flux entre les *clusters* et la cité administrative (A).

Comme discuté dans la section 3.4.4, les domiciles situés à moins de deux kilomètres de la cité administrative forment un seul *cluster* représentant en quelque sorte le centre ville. Cette hypothèse a été faite pour résoudre un problème de lisibilité. En effet, lorsqu'elle n'est pas vérifiée, la *flow map* peut être plus compliquée à analyser.

La figure 3.10 montre un exemple de *flow map* qui se base sur des *clusters* générés sans tenir compte de cette hypothèse, donc avec les domiciles à moins de deux kilomètres de la cité administrative. Un des *clusters* englobe la cité administrative. Donc visuellement parlant, le flux entre ce *cluster* et la cité administrative n'est pas clair car la délimitation entre les deux est floue. Pour cette raison, les domiciles situés à moins de deux kilomètres ne sont pas considérés dans ces flux.

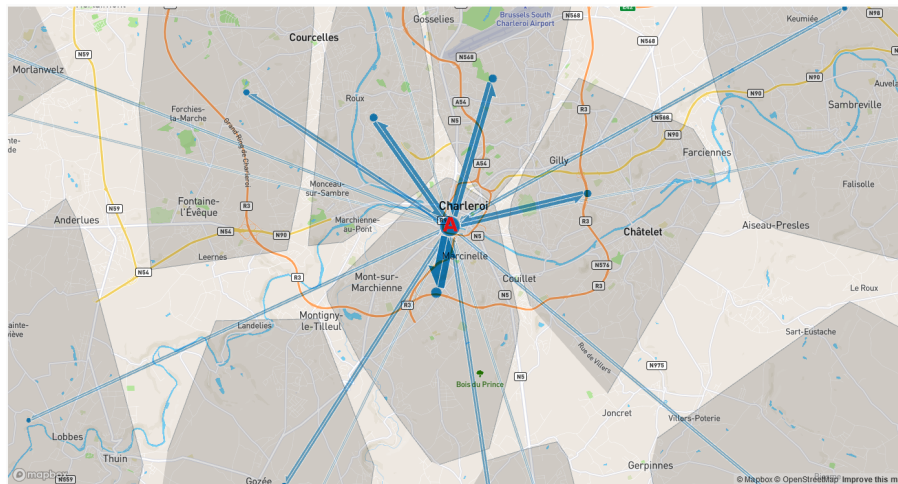


FIGURE 3.10 – Visualisation des flux entre chaque *cluster* et la cité administrative (A) lorsque tous les domiciles sont considérés pour générer les *clusters*. Un de ceux-ci englobe la cité administrative et rend l'analyse de la visualisation plus compliquée.

La technique d'interaction de *filtering* a été implémentée pour permettre à l'utilisateur de choisir quels flux visualiser en fonction de leur magnitude. Ceci permet à l'utilisateur de concentrer son attention sur les flux les plus importants. La figure 3.11 présente la visualisation des flux venant des cinq *clusters* comportant le plus de domiciles.

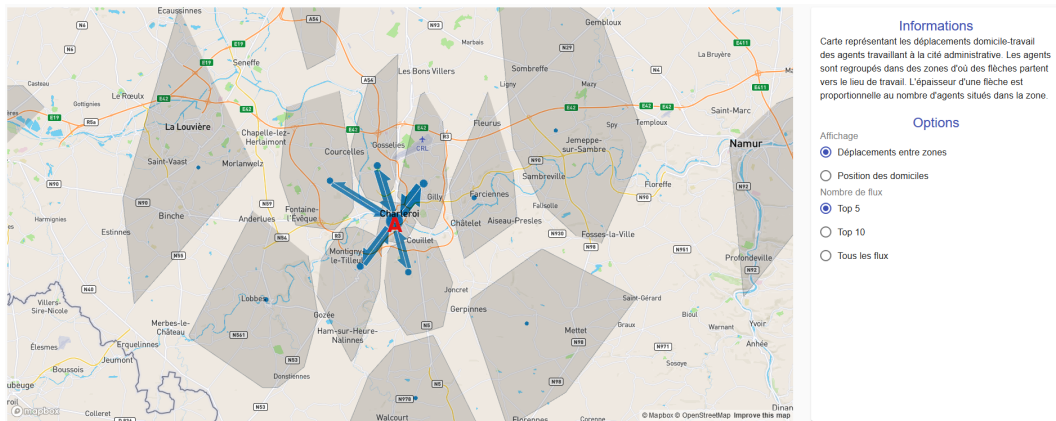


FIGURE 3.11 – Exemple de *filtering* sur la visualisation des flux domicile-travail présentant les cinq *clusters* comportant le plus de domiciles. La cité administrative est représentée par le point A.

Cette visualisation, de par la nature d'une *flow map* ainsi que par la technique d'interaction utilisée, donne une vision macro des déplacements entre les différents domiciles et la cité administrative. En effet, il n'est pas question ici de donner des informations sur les trajets précis empruntés par les agents mais plutôt de donner un aperçu du mouvement global entre les différents *clusters*.

Il est d'ailleurs intéressant de noter que ces *clusters* générés via l'algorithme *k-means* correspondent plutôt bien à des communes. Il y a par exemple un *cluster* positionné vers Gosselies, un autre vers Gilly ou encore un vers Montigny-le-Tilleul. Ceci permet par exemple à l'utilisateur de révéler que la plupart des agents travaillant à la cité administrative habitent au nord de la Ville, vers Gilly, Gosselies ou Courcelles (voir figure 3.11). Cependant cette visualisation ne donne qu'une idée de la position réelle des domiciles. C'est pourquoi une deuxième visualisation a été développée : la visualisation des positions des domiciles, présentée ci-dessous.

La figure 3.12 présente la formalisation de l'analyse WHAT WHY HOW réalisée pour arriver à cette visualisation. La colonne WHAT décrit les données disponibles. Ce sont des *clusters* de points définis au préalable et qui ne changent pas à travers le temps. Ce sont donc des données de type *geometry spatial static* qui sont *categorical* car chaque point fait partie d'un *cluster*. Des flux peuvent être définis entre ces *clusters* et la cité administrative. Ces flux représentent le nombre de points présents dans un *cluster*. Ils sont donc *ordered quantitative sequential*.

La colonne WHY formalise les besoins de la Ville. Celle-ci aimerait pouvoir découvrir des *patterns*, comparer les flux entre eux et rechercher ces flux ainsi que les *clusters* sur la carte.

La colonne HOW est divisée en trois parties. La première, *Encode*, définit les flux et *clusters* sur une carte. Les flux sont représentés avec des flèches de différentes taille en fonction de leur magnitude. Ceci permet à l'utilisateur de comparer les flux entre eux. La deuxième, *Manipulate*, permet à l'utilisateur de sélectionner un *cluster* en passant sa souris par-dessus ainsi que de naviguer sur la carte comme bon lui semble. Enfin, la partie *Reduce* explique qu'il est possible pour l'utilisateur de filtrer les données qu'il affiche. Ceci lui permet de ne visualiser que les flux les plus intéressants et donc de découvrir d'éventuels *patterns*.



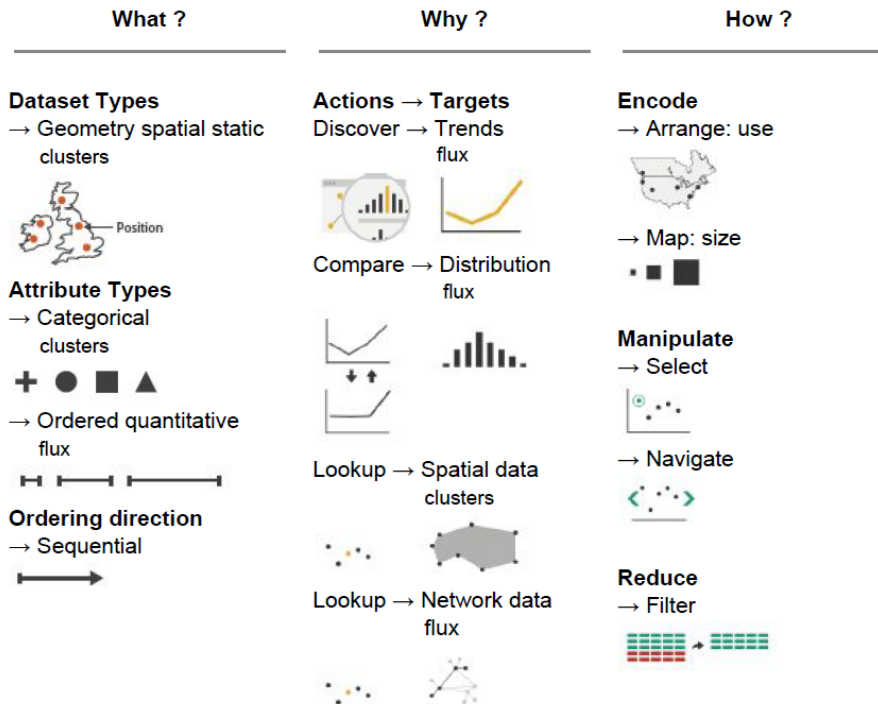


FIGURE 3.12 – Tableau représentant l’analyse WHAT WHY HOW qui a mené à l’élaboration de la *flow map*, présentée à la figure 3.8, qui visualise les flux entre chaque *cluster* et la cité administrative.

**Visualisation des positions des domiciles** La deuxième visualisation à destination des experts a pour objectif de donner des informations sur la position géographique de l’ensemble des domiciles des agents. Ces positions géographiques sont des données spatiales et plus particulièrement des points. La visualisation doit donc utiliser une technique *point-based*. Cette technique est la *heatmap* car elle donne une vision claire de la répartition géographique des domiciles. De plus, elle fournit une information sur la densité des domiciles.

La visualisation est représentée à la figure 3.13. Plus les points sont densément répartis et plus la couleur de la zone concernée tend vers le rouge. A l’inverse, moins les points sont densément répartis, plus la couleur tend vers le jaune. Cette *heatmap* a donc une mesure *ratio* car il est possible de comparer les différentes densités entre elles en terme de ratio d’une densité par rapport à une autre.

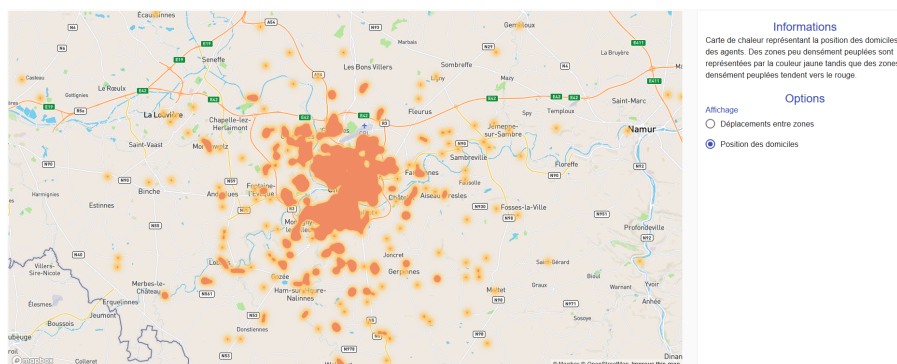


FIGURE 3.13 – Visualisation de la position des domiciles des agents sous la forme d’une *heatmap*. Plus les points sont densément répartis et plus la couleur de la zone concernée tend vers le rouge. Moins les points sont densément répartis et plus la couleur tend vers le jaune.

La figure 3.14 présente la formalisation de l’analyse WHAT WHY HOW faite pour cette *heatmap*. Les données disponibles, la colonne WHAT, sont des points qui symbolisent les adresses des domiciles des agents. Ces points sont du type *Geometry spatial static* et sont ordonnés dans l’espace. Ils ont donc un attribut *ordered quantitative sequential*.

Les actions demandées par la Ville, la colonne WHY, sont les suivantes : découvrir des tendances sur la position des habitations et pouvoir chercher et trouver des emplacements d’habitation sur la carte.

Enfin, la colonne HOW est divisée en plusieurs parties. La première, *Encode*, explique que les points sont encodés sur une carte. Chaque point a une taille et une couleur qui sont fonction de la densité des habitations. Ceci permet à l’utilisateur de découvrir des tendances sur la position des habitations. Ensuite, la partie *Manipulate* explique que l’utilisateur peut naviguer comme il le souhaite sur la carte, ce qui lui permet de chercher et trouver des emplacements d’habitation sur la carte.

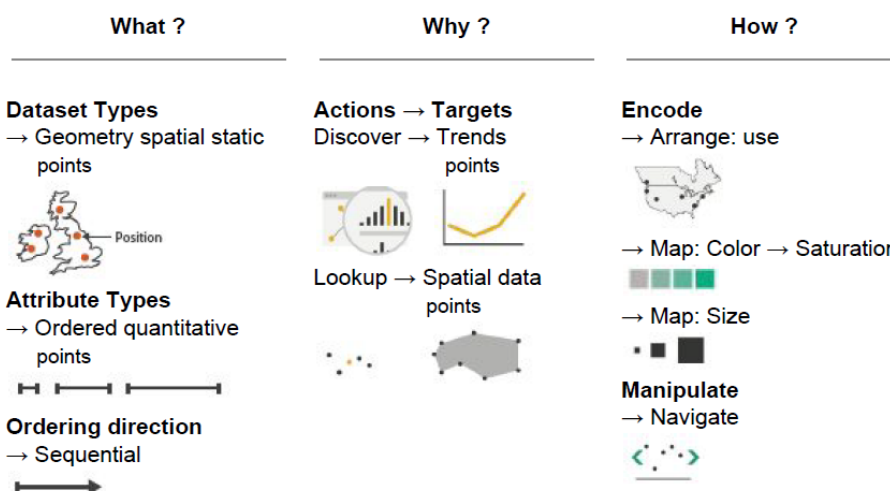


FIGURE 3.14 – Tableau représentant l’analyse WHAT WHY HOW qui a mené à l’élaboration de la *heatmap*, présentée à la figure 3.13, qui visualise la position des domiciles des agents.

**Visualisation des trajets potentiels de l’ensemble des agents** Cette visualisation a pour objectif de donner des informations plus précises sur les déplacements des agents entre leur domicile et leur endroit de travail et ce, en fonction du moyen de transport utilisé. Pour rappel, le réseau généré dans la section 3.4.4 est constitué d’un ensemble de segments de route ayant chacun un poids représentant le nombre de fois que ce segment est emprunté. Cet ensemble de données est donc composé de lignes. En toute logique, la technique utilisée sera de la famille *line-based*.

Parmi toutes les techniques *line-based*, la *flow map* est intéressante. En effet, elle permet, dans ce cas-ci, de représenter les trajectoires des agents sans perte d’information car chacune d’elle est présente sur la carte. De plus, la représentation des poids de chaque segment sur la *flow map* permet de révéler de possibles *patterns* et de focaliser l’attention de l’utilisateur sur les segments les plus utilisés. Pour toutes ces raisons, cette technique de visualisation *line-based* a été implémentée.

Afin de répondre à l’objectif de cette visualisation, il est nécessaire de représenter les informations disponibles sur le trajet aller entre le domicile et le travail ainsi que sur le trajet retour entre le travail et le domicile. Une première idée fut de définir des trajets pour chaque *cluster* comme à la figure 3.15. Cependant il était plus intéressant pour la Ville d’avoir des informations plus précises.



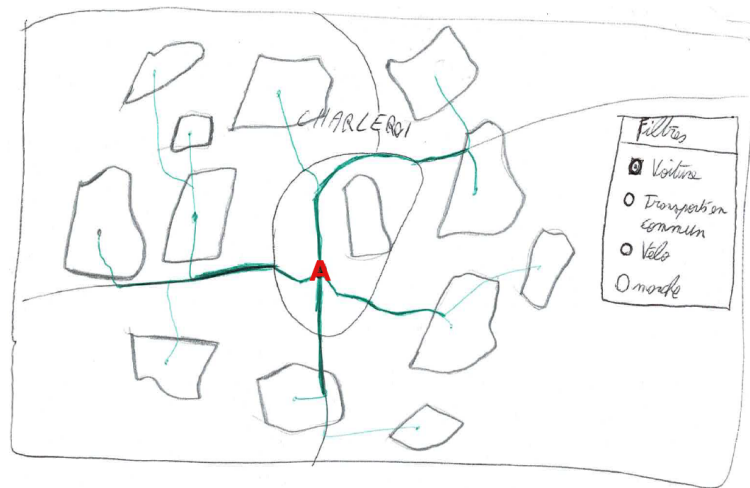


FIGURE 3.15 – *Mockup* présentant une idée de *flow map* visualisant des propositions de trajets entre les *clusters* et la cité administrative (A). Les segments verts représentent les trajets. Plus un segment est épais et plus son poids est élevé. Cette idée fut écartée pour privilégier la visualisation de trajets depuis tous les domiciles.

La figure 3.16 présente donc la visualisation finale. Elle affiche les potentiels trajets effectués en voiture par l'ensemble des agents dont le domicile se trouve à plus de 5 km de la cité administrative. La carte de gauche représente les trajets effectués le matin pour se rendre au travail et la carte de droite les trajets effectués le soir pour retourner à son domicile.

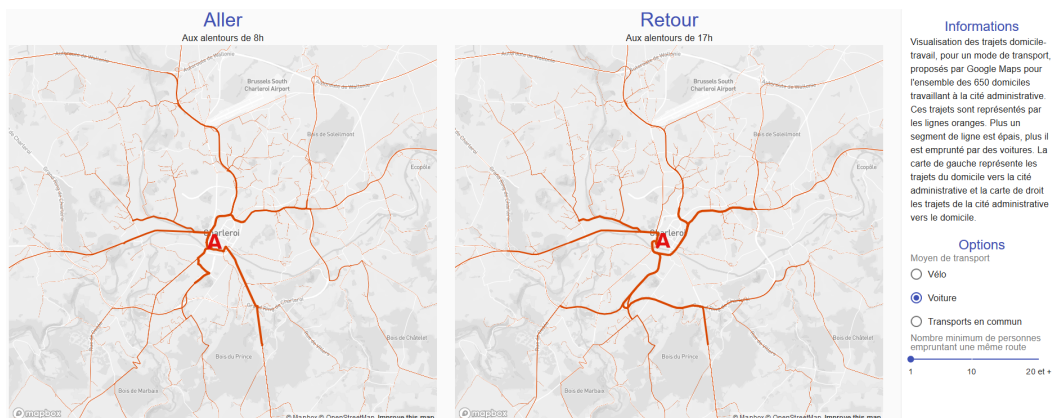


FIGURE 3.16 – Visualisation des trajets effectués en voiture par les agents habitant à plus de 5 km de la cité administrative (A).

Cette division en deux cartes, une pour se rendre à son lieu de travail et l'autre pour revenir à son domicile, n'est pas appliquée pour les trajets en transport en commun. En effet, les lignes de transport en commun sont les mêmes dans les deux cas ; les trajets allers et retours peuvent donc être agrégés sur une seule carte. Ce cas de figure est présenté à la figure 3.17.

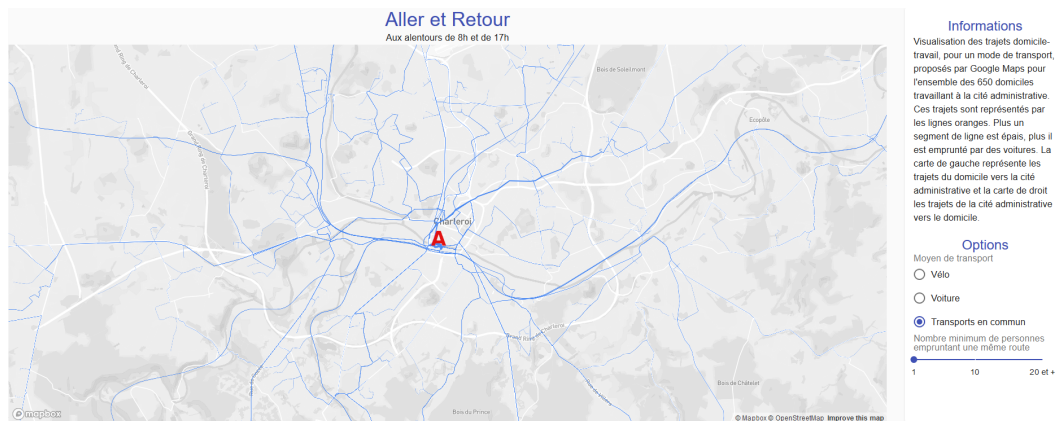


FIGURE 3.17 – Visualisation des trajets effectués en transports en commun par les agents habitant à plus de 5 km de la cité administrative (A).

De la même manière que la vue globale des flux, le *filtering* a été implémenté et permet à l'utilisateur de choisir quel type de trajet visualiser. La figure 3.16 présente les trajets effectués en voiture, mais il est possible de visualiser les trajets en transports en commun et en vélo. Pour plus de clarté, une couleur est attribuée à chaque moyen de transport. Par exemple, la voiture est représentée en rouge et les transports en commun en bleu. De plus, l'utilisateur peut choisir de ne visualiser que les segments de route empruntés par un certain nombre de personnes avec le *slider*. La figure 3.18 montre par exemple uniquement les segments de route empruntés par au moins 10 agents, dans le cas de la voiture.

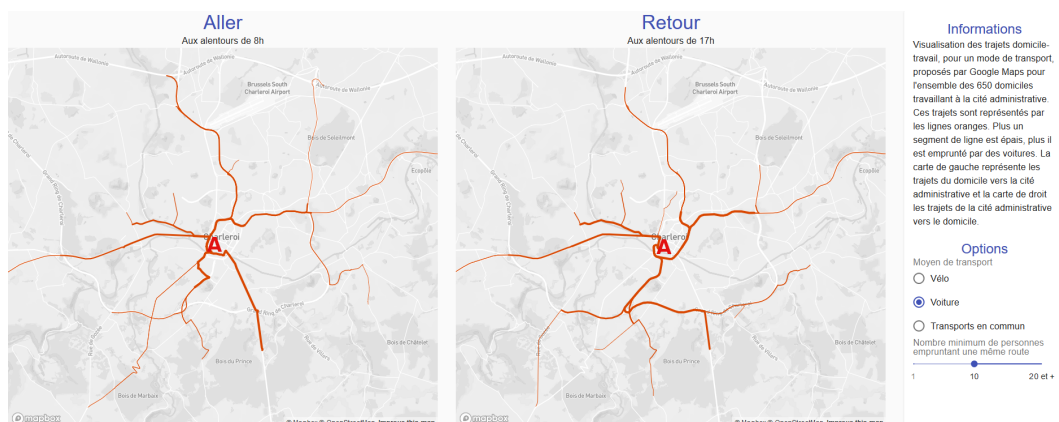


FIGURE 3.18 – Visualisation des trajets effectués par les agents habitant à plus de 5 km de la cité administrative (A) et empruntés par au moins 10 agents.

La figure 3.19 présente la formalisation de l'analyse **WHAT WHY HOW** faite pour définir cette visualisation. Les données, la colonne **WHAT**, sont des trajets constitués d'un ensemble de positions successives. Donc ce sont des données de type *Geometry spatial static*. Chaque segment de route est classé en fonction d'un moyen de transport (*Categorical*) et possède un nombre de trajets (*Ordered quantitative sequential*).

Les actions à réaliser, la colonne **WHY**, sont les suivantes : découvrir des tendances dans les trajets effectués, pouvoir chercher et trouver des trajets sur la carte et savoir comparer les trajets entre eux en fonction du nombre de voitures qui les empruntent.

La colonne **HOW** est divisée en quatre parties. La première, *Encode*, décrit comment les données sont encodées sur la visualisation. Chaque trajet est représenté sur une carte. L'épaisseur d'un trajet est fonction du nombre de voitures empruntant ce

trajet. Ceci permet à l'utilisateur de découvrir des tendances d'utilisation de segments de trajets. La partie *Manipulate* décrit que l'utilisateur peut naviguer sur la carte, ce qui lui permet de chercher et trouver les trajets qui l'intéressent. La partie *Facet* explique comment les données sont présentées. Les trajets aller et retour sont visualisés l'un à côté de l'autre, chacun sur une carte. Cela permet à l'utilisateur de comparer les trajets aller-retour. Enfin, la partie *Reduce* explique qu'il est possible de filtrer les informations présentes sur la visualisation en fonction du moyen de transport et du nombre de véhicules, ce qui aide l'utilisateur à comparer les trajets entre eux.

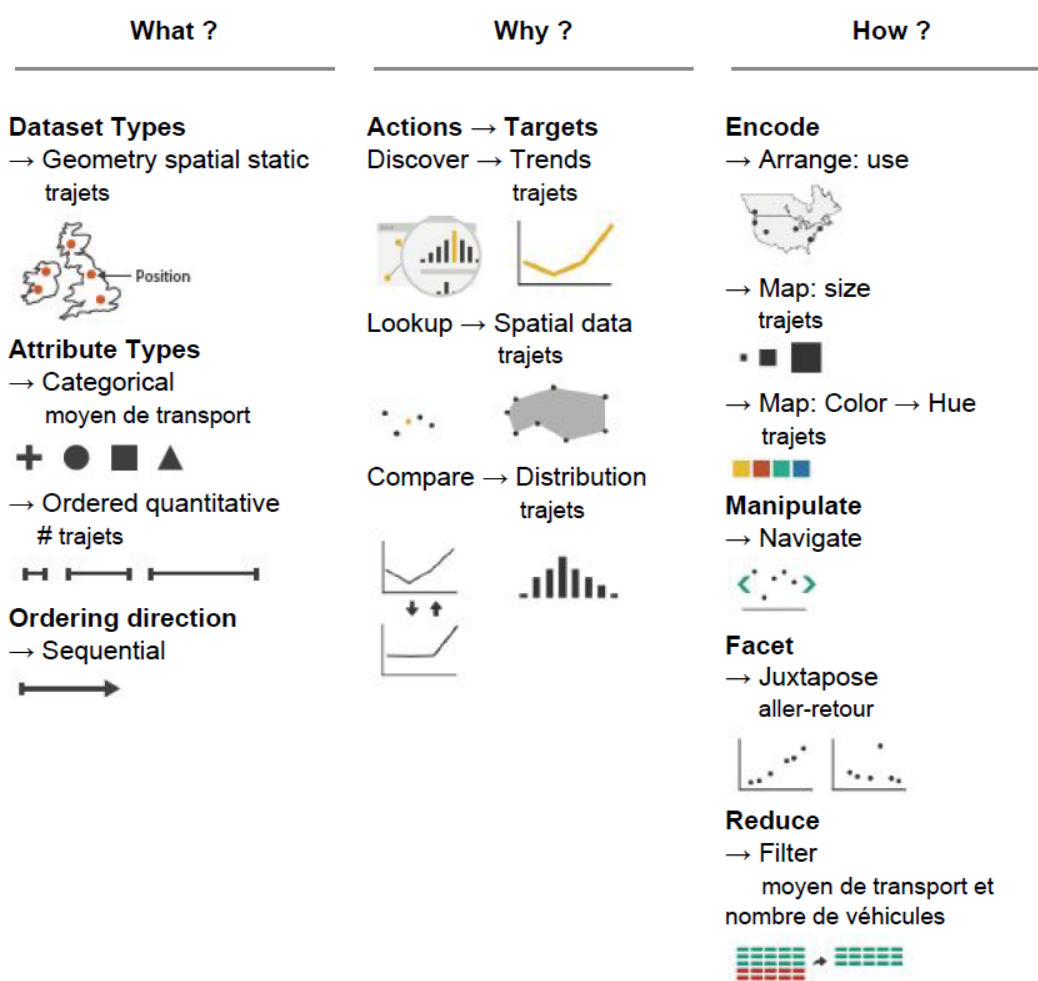


FIGURE 3.19 – Tableau représentant l'analyse WHAT WHY HOW qui a mené à l'élaboration de la *flow map* à destination de la Ville, présentée à la figure 3.16, et qui visualise les trajets effectués par les agents en fonction de leur moyen de transport et de la distance les séparant de la cité administrative.

Le développement de cette plateforme a rapidement été mis de côté au profit de la plateforme à destination de l'ensemble des agents. En effet, les différentes réunions organisées avec les clients ont permis de réaliser que la plateforme n'allait pas être utilisée telle quelle par la Ville. Ceci est discuté en détail dans la section 3.4.7.

**Visualisation des trajets proposés pour un agent** Cette visualisation est développée pour remplir plusieurs objectifs. Tout d'abord, elle doit informer chacun des agents qui l'utilisent sur les différents moyens de transport possibles entre domicile et lieu de travail. Ensuite, elle doit pousser ces mêmes agents à privilégier les transports durables comme la marche, le vélo et les transports en commun, au détriment de l'utilisation de la voiture personnelle. La visualisation doit donc permettre d'avoir une vision des

trajets proposés en fonction des différents moyens de transport ainsi que d'avoir des informations précises sur chacun de ces trajets afin de pouvoir les comparer.

Une première idée d'interface est montrée dans la figure 3.20. Cependant elle fut abandonnée car trop complexe. La partie 1 n'apporte aucune plus-value et la partie 2 peut être simplifiée en une seule carte. A partir de ces constats, la visualisation finale a pu être développée et est présentée ci-dessous. Après le développement de cette plateforme, un mode d'emploi a été rédigé afin de simplifier au maximum la prise en main de l'outil.

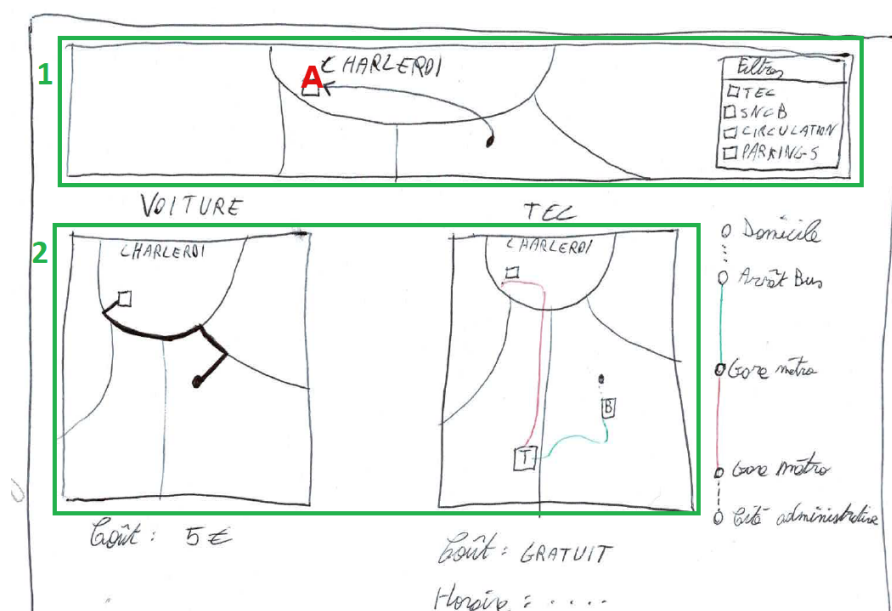


FIGURE 3.20 – Mockup d'une interface de la visualisation à destination des agents. Il présente tout d'abord la carte générale (1) montrant le flux entre le domicile et la cité administrative (A). Il présente ensuite les différents trajets (2) et les compare. Cette interface a été abandonnée car elle est trop complexe pour le peu d'informations qu'elle donne à l'utilisateur.

La visualisation est définie en fonction de la requête de l'utilisateur, elle-même définie en fonction d'un formulaire. Ce formulaire est composé de deux paramètres : l'adresse de son domicile et son lieu de travail, que l'utilisateur choisit parmi tous ceux disponibles sur la plateforme. La figure 3.21 montre l'interface de base proposée lorsque l'utilisateur arrive sur la plateforme. Elle est très épurée et ne présente qu'une carte vierge à gauche et le formulaire que l'utilisateur doit remplir.

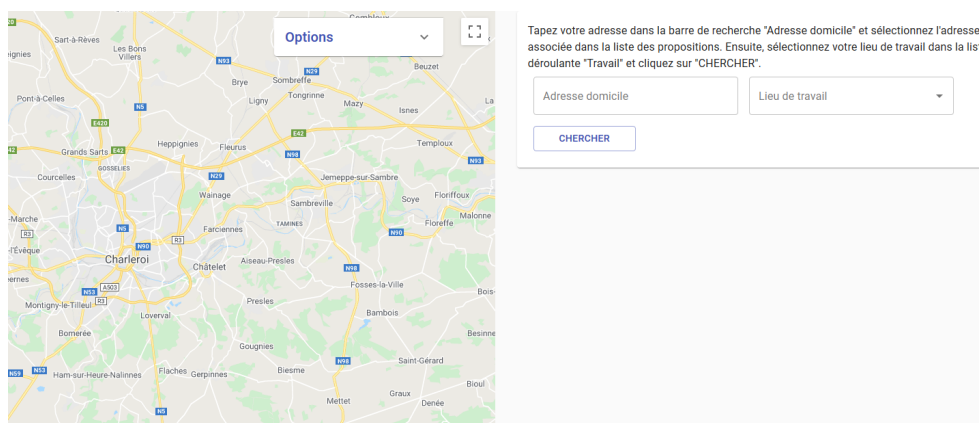


FIGURE 3.21 – Interface proposée lorsque l'utilisateur arrive sur la plateforme. La partie gauche présente une carte Google Maps vierge et la partie droite contient un formulaire.

Une fois la requête envoyée à la plateforme, la visualisation est définie (voir figure 3.22). La partie gauche de la visualisation est composée d'une carte représentant une vision géographique des différents trajets proposés, chacun d'eux représentant un moyen de transport. Dans cette figure, le domicile est situé à 4.2 km à vol d'oiseau du lieu de travail. Par conséquent, les trajets proposés sont le vélo, la voiture et les transports en commun. Elle représente le trajet en vélo en vert, le trajet en voiture en rouge et le trajet en transports en commun en bleu.

La partie droite, quant à elle, affiche les informations importantes de chacun de ces trajets via notamment un tableau comparatif en fonction de différentes variables comme la distance du trajet, sa durée, son coût annuel et sa pollution annuelle. En ce qui concerne le coût, la voiture est désavantagée par rapport aux autres moyens de transport car son coût annuel est totalement à charge de l'agent. En revanche, la Ville rembourse les trajets effectués en vélo à hauteur d'une somme forfaitaire au kilomètre et rembourse totalement les abonnements de transports en commun.

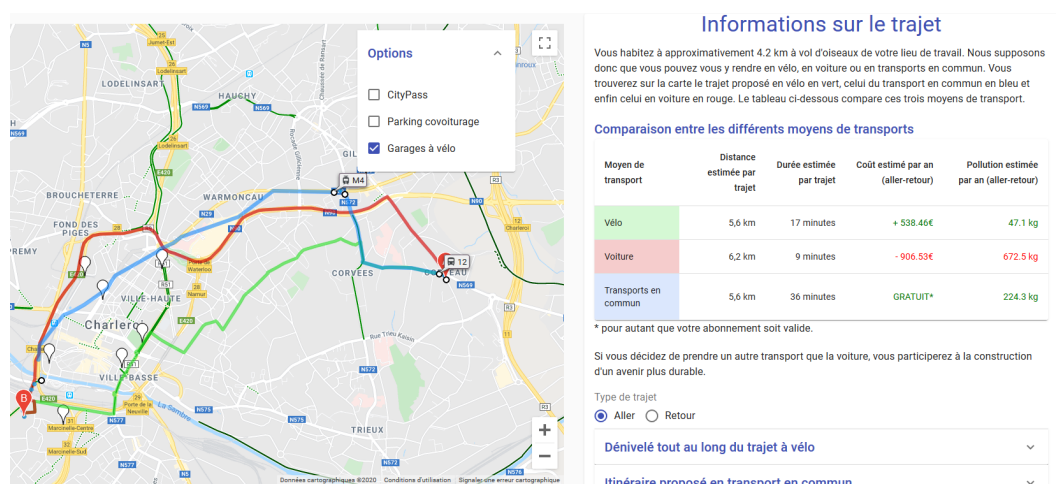


FIGURE 3.22 – Visualisation des trajets proposés pour une adresse d'un agent travaillant à l'Ecologie Urbaine de la Ville. La partie gauche reprend la carte avec les différents trajets proposés en fonction des moyens de transport (vélo, voiture et transports en commun). La partie droite reprend les informations précises sur les trajets proposés.

La visualisation permet aussi à l'agent d'interagir de trois manières différentes. La première est la visualisation de couches de données statiques par-dessus la carte. Ces couches sont des données KML, discutées dans la section 3.2.7, regroupées dans l'encadrement des options, en haut à droite de la carte. Cet encadrement agit comme un système de filtres permettant d'afficher ou non ces couches de données. Par exemple, toujours sur la même figure, la couche "Garages à vélo" est représentée par l'ensemble des marqueurs blancs positionnés sur la carte, chaque marqueur représentant un garage à vélo mis à disposition par la Ville. Cette technique de *filtering* permet à l'utilisateur de choisir de ne voir que les données qui sont intéressantes pour lui tout en ne chargeant pas trop la visualisation.

La deuxième est toujours une technique de *filtering* permettant ici de montrer les différents trajets en fonction de si ceux-ci sont les trajets pour aller à son lieu de travail ou pour retourner à son domicile. Ce filtre s'applique à la carte mais aussi aux informations plus précises sur les trajets proposés. Pour l'aller, les trajets sont générés pour que l'agent arrive à maximum 8h30 à son lieu de travail. Pour le retour, les trajets sont générés en partant du principe que l'agent quitte son travail à 17h. Lorsqu'un utilisateur effectue une recherche sur la plateforme, l'ensemble des trajets proposés, aller et retour, sont générés pour le lendemain. C'est-à-dire que si, par exemple, l'agent fait sa recherche le dimanche 26 janvier, les résultats montrés sont les prédictions faites par Google Maps pour le lundi 27 janvier.



La troisième concerne des informations plus précises. En effet, si l'agent le souhaite, il peut choisir d'afficher des informations précises sur les trajets proposés. Ces informations sont de différentes sortes en fonction du moyen de transport. En ce qui concerne les trajets proposés en vélo, un graphique présentant le dénivelé tout au long du trajet est disponible afin d'aider les agents qui n'ont pas l'habitude de ce mode de transport à avoir une idée de la difficulté du trajet proposé. De plus, afin d'aider l'utilisateur à visualiser les endroits difficiles, le graphique est lié au trajet correspondant sur la carte. Si l'utilisateur survole un point du graphique, le point correspondant du trajet est visualisé sur la carte via un marqueur Google Maps.

Un exemple est montré à la figure 3.23. En ce qui concerne les transports en commun, un horaire est proposé en fonction du type de trajet sélectionné par l'agent, aller ou retour. Un lien vers le site de la TEC ou la SNCB est proposé aux utilisateurs qui désireraient faire des recherches sur d'autres horaires. La figure 3.24 montre un exemple d'itinéraire de transport en commun proposé pour effectuer le trajet aller présenté à la figure 3.22.

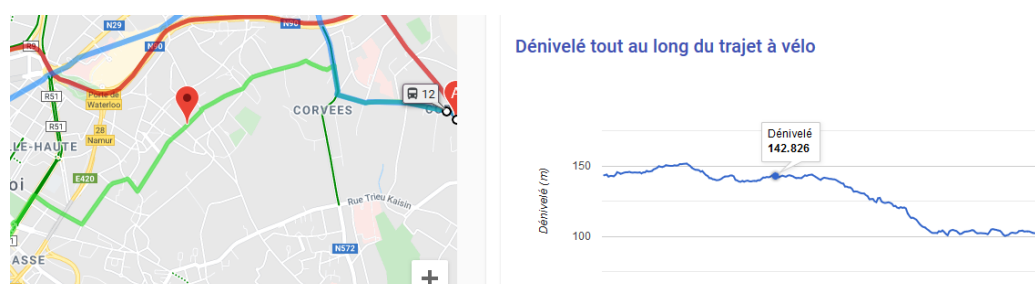


FIGURE 3.23 – Exemple de visualisation du dénivelé d'un trajet en vélo. Chaque point du graphique est lié à son point correspondant sur le trajet visualisé sur la carte.

#### Itinéraire proposé en transport en commun

Le calcul d'itinéraires piétons est en bêta. Faites attention – Cet itinéraire n'est peut-être pas complètement aménagé pour les piétons.

**A** Rue Lison 170, 6200 Charleroi, Belgique

5,6 km. Environ 36 minutes

Marcher jusqu'à GILLY Corvées  
Environ 1 minute

**GILLY Corvées**

12 Bus vers GILLY Cimetière  
07:51–07:56 (5 minutes, 4 arrêts)  
Service assuré par [TEC Charleroi](#)

**GILLY Gazomètre (Ch. de Montignies)**

Marcher jusqu'à GILLY Gazomètre (M)  
Environ 2 minutes

**GILLY Gazomètre (M)**

M4 Tramway vers CHARLEROI Sud  
08:10–08:20 (10 minutes, 5 arrêts)  
Service assuré par [TEC Charleroi](#)

**CHARLEROI Villette (M)**

Marcher jusqu'à Rue Julien Dulait 22, 6001 Charleroi, Belgique  
Environ 6 minutes

**B** Rue Julien Dulait 22, 6001 Charleroi, Belgique

FIGURE 3.24 – Itinéraire proposé en transport en commun pour effectuer le trajet présenté à la figure 3.22.

La figure 3.25 présente la formalisation de l'analyse WHAT WHY HOW réalisée pour définir cette visualisation. Les données, la colonne WHAT, proviennent des API Google et des KML de Google MyMaps. Il y a donc des données de type *Geometry spatial dynamic* pour les trajets et les données KML, des *Tables* pour les détails sur chaque trajet et un *Fields* pour le dénivelé. En effet, le dénivelé est une donnée continue contrairement aux autres. Les trajets sont classés par moyen de transport et les données provenant des KML sont elles aussi classées. Ces deux types de données ont donc un attribut *Categorical*.

Les actions demandées par la Ville, la colonne WHY, sont multiples. La Ville aimerait que les trajets et les détails des trajets soient présentés à l'utilisateur afin que celui-ci puisse les comparer. Enfin, elle aimerait que les détails des trajets soient résumés afin de ne pas noyer l'utilisateur dans trop de données.

Pour faire tout cela, la visualisation est définie dans la colonne HOW. La partie *Encode* décrit que les trajets sont encodés sur une carte et les détails des trajets sont présentés dans un tableau. Chaque trajet a une couleur en fonction de son moyen de transport et les données KML sont affichées avec différents marqueurs Google Maps. Ceci permet de présenter des données résumées à l'utilisateur. La partie *Manipulate* décrit les différentes manipulations que l'utilisateur peut faire sur la visualisation. Il peut sélectionner un trajet et naviguer sur la carte, ce qui peut l'aider à comparer les différents trajets. La partie *Facet* explique que les trajets sont présentés sur une seule carte, les uns sur les autres. Enfin, la partie *Reduce* explique qu'il est possible de filtrer les informations présentes sur la carte en donnant à l'utilisateur la possibilité de choisir quelles couches de données KML il choisit de visualiser.

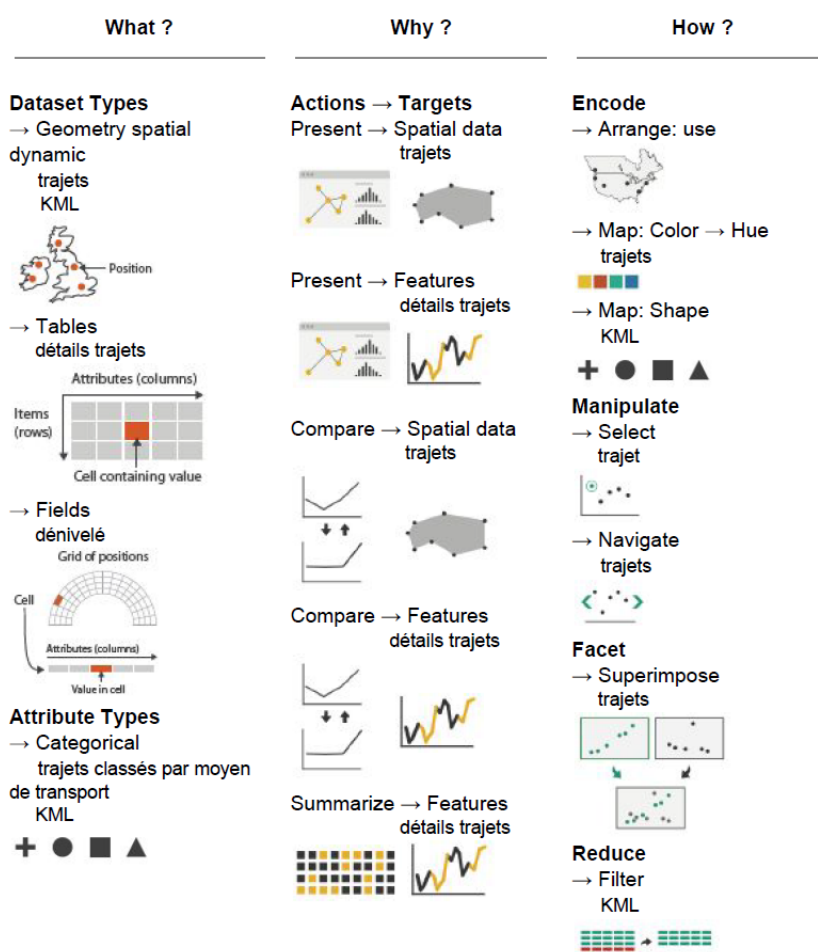


FIGURE 3.25 – Tableau représentant l'analyse WHAT WHY HOW qui a mené à l'élaboration de la visualisation à destination des agents présentée à la figure 3.22.

### 3.4.6 Coût de la solution proposée

Comme la plateforme à destination des experts a été mise de côté, le coût de la solution discuté ici concerne uniquement la plateforme à destination des agents de la Ville. Comme mentionné dans la section 3.4.2, l'utilisation des API de Google Maps implique un certain coût pour la Ville. En effet, Google Maps fait payer l'utilisation de ses API en fonction du nombre de requêtes effectuées.

Cette section calcule une estimation du coût de recherche d'une adresse sur la plateforme, c'est-à-dire combien cela coûte à la Ville pour chaque recherche effectuée par un agent sur la plateforme. Les prix sont donnés sur le site de Google Maps Platform en dollars. Cependant les factures sont calculées en euros. La suite du raisonnement est donc en euros. La facture de l'utilisation des API de Google Maps au cours du mois d'avril, présente à la figure 3.26, permet de calculer le coût d'envoi de 1000 requêtes à chacune des API.

Produit	Type de ressource	Intervalle	Utilisation	Montant (€)
Directions API	Directions	1 avr. - 30 avr.	8296 unités	37,70
Places API	Query Autocomplete - Per Request	1 avr. - 30 avr.	11457 unités	29,47
Places API	Places - Text Search	1 avr. - 30 avr.	505 unités	14,69
Directions API	Directions Advanced	1 avr. - 30 avr.	1231 unités	11,19
Places API	Places Details	1 avr. - 30 avr.	629 unités	9,72
Maps API	Dynamic Maps	1 avr. - 30 avr.	1359 unités	8,65
Places API	Atmosphere Data	1 avr. - 30 avr.	1134 unités	5,15
Maps Elevation API	Elevation	1 avr. - 30 avr.	856 unités	3,89
Places API	Contact Data	1 avr. - 30 avr.	1134 unités	3,09
Geocoding API	Geocoding	1 avr. - 30 avr.	470 unités	2,14
Credit	Maps Free Tier	1 avr. - 30 avr.		-125,69
Sous-total en EUR				0,00 €
TVA (21%)				0,00 €
Total en EUR				<b>0,00 €</b>

FIGURE 3.26 – Facture d'utilisation des API Google Maps durant le mois d'avril 2020. Le coût total après remise est nul car Google Maps offre 200\$ de crédit par mois sur l'utilisation de ses API.

Plusieurs détails de cette facture méritent une attention particulière. Tout d'abord, le coût total final est de 0€. En effet, Google Maps offre 200\$, soit environ 183€<sup>3</sup>, de crédit par mois sur l'utilisation de ses API. C'est un point à prendre en compte lors de la budgétisation du coût total de la plateforme. Ensuite, certaines API discutées dans la section 3.4.2 sont divisées en plusieurs sous-API, chacune ayant un coût différent. Pour plus de simplicité, ces divisions sont rassemblées pour calculer un coût général pour chaque API<sup>4</sup> :

1. *Maps API* : 6,36€ pour 1000 requêtes
2. *Geocoding API* : 4,55€ pour 1000 requêtes
3. *Directions API* : 5,13€ pour 1000 requêtes
4. *Elevation API* : 4,55€ pour 1000 requêtes
5. *Places API* : 4,18€ pour 1000 requêtes

Le coût moyen de l'ensemble de ces API dépend donc du nombre de requêtes envoyées vers chacune d'elles. Il est calculé en divisant le coût total avant remise par le nombre total

3. La conversion entre dollar et euro a été faite le 2 mai 2020 via le taux de change entre ces deux monnaies fourni par l'API de la Banque Centrale Européenne via ce lien <https://api.exchangeratesapi.io/latest>.

4. L'API *Charts* est gratuite. Elle n'est donc pas mentionnée dans la suite du calcul.



de requêtes envoyées vers toutes les API et donne une valeur de 4,64€ pour 1000 requêtes. Ce coût moyen est une estimation du coût réel. Cela signifie concrètement que lorsque la plateforme effectue 1000 requêtes sur les API Google Maps, cela coûte environ 4,64€ à la Ville.

Pour planifier ses budgets, la ville de Charleroi a besoin de savoir combien la plateforme va lui coûter et donc combien de requêtes vont être envoyées. Pour ce faire, il est nécessaire de calculer le nombre de requêtes envoyées aux API pour une recherche effectuée sur la plateforme par un agent. Ce nombre dépend de la distance séparant le domicile de l'agent de son lieu de travail. Si elle est inférieure à 2 km, deux moyens de transport sont proposés : la marche et le vélo. Si la distance est supérieure à 2 km et inférieure à 5 km, trois moyens de transport sont proposés : le vélo, la voiture et les transports en commun. En revanche, si elle est supérieure à 5 km, il n'y a que deux moyens de transport proposés : la voiture et les transports en commun.

Afin de calculer l'estimation la plus large, il est nécessaire de calculer le nombre de requêtes envoyées aux API dans le cas où ce nombre est le plus grand. Ce cas est celui pour lequel la distance est comprise entre 2 km et 5 km car il propose plus de moyens de transport et nécessite donc plus d'appels aux API. Voici les requêtes envoyées aux API par la plateforme dans ce cas de figure :

1. L'utilisateur écrit son adresse et sélectionne une proposition parmi la liste : environ 18 requêtes
2. La plateforme calcule les trois trajets aller et retour : maximum 8 requêtes
3. La plateforme calcule le dénivelé des trajets aller et retour du vélo : 2 requêtes

Lorsqu'un utilisateur lance une recherche sur la plateforme, il n'y a donc normalement pas plus de 28 requêtes envoyées aux API. Pour plus de sécurité, ce nombre maximum est revu à la hausse et passe à 30 requêtes par recherche. Grâce au coût moyen pour 1000 requêtes estimé ci-dessus, le coût de recherche sur la plateforme est donc d'environ 0,14€. Cela veut dire que pour chaque recherche effectuée par un agent sur la plateforme, la Ville doit payer environ 0,14€ à Google.

### 3.4.7 Évaluation de la solution proposée

Tout au long du développement, il est primordial de réaliser une évaluation utilisateur de la solution proposée. Celle-ci permet de mettre en avant les points forts de la solution ainsi que ses limites. La moitié du stage s'est déroulée durant le confinement lié au Covid-19. Dans ce contexte, réaliser une évaluation du produit par l'utilisateur est compliqué. En effet, l'utilisateur ne peut pas tester lui-même le produit et ne peut l'évaluer qu'à travers les démonstrations faites par le développeur par vidéo-conférence. De ce fait, l'évaluation utilisateur s'est principalement faite via les retours donnés par le client lors des réunions hebdomadaires ou bi-mensuelles. Cette évaluation a permis de révéler les points forts et les limites de la solution.

**Points forts de la solution** Les points forts discutés sont ceux de la visualisation à destination des agents de la Ville. En effet, comme vu ci-dessus, les visualisations à destination des experts ont été mises de côté et ne sont donc pas terminées. Dès lors, il n'est pas pertinent de discuter de leurs points forts.

Un point qui est revenu très souvent est la facilité d'utilisation de l'application. Cette facilité provient de la conjonction de deux facteurs : l'interface et le nombre restreint de paramètres gérés par l'utilisateur.

En ce qui concerne le premier facteur, lorsque l'utilisateur arrive sur la plateforme, l'interface proposée est épurée et très simple comme montrée dans la figure 3.21. A gauche se trouve une carte Google Maps vierge centrée sur Charleroi. A droite se trouve un formulaire permettant à l'utilisateur d'entrer son adresse et de sélectionner son lieu de travail. C'est tout ce que propose cette interface de base. Lorsque l'utilisateur effectue sa recherche, l'interface change et les informations sont visualisées comme dans la figure 3.22. Ce choix de ne pas surcharger l'utilisateur d'informations lorsque celui-ci

arrive sur la plateforme a pour but de lui faciliter au maximum la prise en main de l'application.

Le deuxième facteur concerne l'interaction restreinte entre l'utilisateur et la plateforme. Il n'est pas question ici de l'interaction entre l'utilisateur et la visualisation une fois que les données sont visualisées mais bien de l'interaction nécessaire entre l'utilisateur et la plateforme avant que les données ne soient générées. Cette interaction est représentée par le formulaire présent dans la figure 3.21. L'utilisateur doit simplement rentrer l'adresse de son domicile et de spécifier où il travaille. La plateforme s'occupe du reste.

Un deuxième point fort de l'application est qu'elle rassemble en un seul endroit plusieurs données et informations différentes afin de créer de la nouvelle information à destination de l'utilisateur. Ces données existent déjà pour la plupart mais elles proviennent de différentes sources qui n'ont à l'origine aucun lien entre elles. Voici une liste exhaustive des données regroupées en un seul point par la plateforme :

1. L'ensemble des données fournies par les API Google Maps et discutées dans la section 3.4.2 sont disponibles sur Google Maps à un niveau individuel.
2. Les données KML discutées dans la section 3.2.7.
3. Les différents lieux de travail des agents discutés dans la section 3.2.6.
4. Les constantes définies par la Ville permettant de calculer le coût d'utilisation annuel de la voiture ainsi que le remboursement annuel des trajets en vélo
5. Les constantes de pollution par km en fonction du moyen de transport et définies par la Wallonie.

Le rassemblement de ces données donne un caractère multi-fonction à l'application qui est fort apprécié par le client.

**Limites de la solution** Concernant les visualisations à destination des experts, les réunions ont permis de se rendre compte rapidement des limites de cette première partie de solution. Tout d'abord, les personnes travaillant dans le domaine de la mobilité réalisent des analyses de cartes presque exclusivement sur QGIS. De ce fait, une plateforme web comme celle proposée ici ne s'intègre pas bien dans leurs habitudes. Ensuite, la plateforme permet de réaliser des analyses poussées des déplacements des 650 agents travaillant à la cité administrative mais pas de l'ensemble des 3500 agents de la Ville. S'il était possible de réaliser les analyses de déplacement faites sur la cité administrative pour l'ensemble des bâtiments dans lesquels des agents travaillent, la plateforme pourrait potentiellement être plus utilisée.

Pour toutes ces raisons, le développement des visualisations à destination de la Ville a été rapidement mis de côté au profit de la visualisation à destination des agents. Une conséquence est que ces deux visualisations fournissent une analyse qui pourrait être affinée. Ceci est discuté plus en détails dans le chapitre 4.

Concernant la visualisation à destination des agents, l'évaluation avec les utilisateurs s'est globalement faite de la même manière que les visualisations à destination des experts. Cependant, une démonstration a été faite devant l'ensemble des membres du cabinet de Mobilité de la Ville afin d'avoir leur avis sur l'application, lorsque le développement fut presque terminé. L'avantage de cette démonstration est que les membres du cabinet sont des utilisateurs potentiels de l'application. Les limites présentées ci-dessous proviennent principalement des réunions avec les clients mais aussi des retours des membres du cabinet lors de cette démonstration.

La limite la plus importante est que l'application proposée n'est pas facilement paramétrable par la Ville. Pendant le développement, plusieurs choix ont été faits en concertation avec la Ville et ceux-ci peuvent influencer les résultats générés par la plateforme. Voici ces choix :

1. Les premiers choix sont la distance à vol d'oiseau séparant le domicile du lieu de travail et les trajets proposés. Ces deux choix sont intimement liés. Lorsqu'une certaine distance sépare le domicile de l'utilisateur de son lieu de travail, des moyens de transport adaptés sont proposés. Pour le moment, si cette distance est inférieure à 2 km, la marche et le vélo sont mis en avant. Si elle est entre 2 km et 5 km, le vélo, la voiture et les transports en commun sont suggérés. Si elle est supérieure à 5 km, la voiture et les transports en commun sont présentés.

Il est possible d'agir sur ces deux choix. Soit les différentes distances décrites ci-dessus sont fixées, il est alors possible de modifier les transports pour chacune de ces distances et par exemple proposer la voiture en plus de la marche et du vélo pour une distance inférieure à 2 km. Soit les moyens de transports avancés sont fixés, il est alors possible de modifier les distances correspondantes pour par exemple favoriser le vélo et la marche pour une distance inférieure à 3 km au lieu de 2 km. Soit les distances et les choix proposés peuvent être modifiés de concert en suggérant par exemple la voiture, le vélo et la marche pour une distance inférieure à 3 km.

2. Les coûts et la pollution sont calculés au km via des constantes. Par exemple, le vélo est remboursé à hauteur de 0,24€ au km.
3. Certains lieux de travail se voient attribuer un malus de temps à ajouter au temps estimé de trajet pour les voitures. Ce malus symbolise le temps nécessaire pour se parquer.
4. De la même manière que le point précédent, certains lieux de travail se voient attribuer un malus de 120€ de coût annuel estimé de la voiture. Ce malus symbolise un abonnement de parking imposé aux personnes voulant se parquer dans l'intra-ring. Ce malus est donc ajouté aux lieux de travail se situant dans l'intra-ring.

Malheureusement ces choix ne sont pas paramétrables par la Ville. Si celle-ci veut changer ces paramètres pour une raison ou une autre, elle ne peut le faire qu'en modifiant directement les valeurs dans le code de l'application ce qui n'est pas pratique voire très difficilement faisable pour des non-informaticiens.

Dans le même ordre d'idée, la plateforme n'est également pas paramétrable pour l'utilisateur. En effet, si celui-ci souhaite par exemple que la plateforme lui propose un trajet en vélo alors qu'il habite à plus de 5 km de son lieu de travail, ce n'est pas possible.

### 3.5 Discussion sur les contributions

Le stage effectué à Charleroi ainsi que ce mémoire proposent diverses contributions qui se répartissent en deux catégories : une contribution théorique et des contributions pratiques. Ces deux types de contributions sont discutés ci-dessous.

La contribution théorique est la réponse à la question de recherche de ce mémoire. Pour ce faire, un état de l'art des techniques de visualisation de trafic a été réalisé. Cet état de l'art fournit les outils qui ont permis de construire une proposition de solution. Cette solution est divisée en deux afin de répondre aux deux parties de la question de recherche, c'est-à-dire d'une part, connaître les moyens de déplacement domicile-travail des agents et d'autre part, proposer à ces agents différents moyens de transport pour effectuer ces mêmes trajets de la manière la plus durable possible.

Suite à la rédaction de l'état de l'art et à la définition de la solution, une question se pose. L'état de l'art et la solution répondent-ils entièrement à la question de recherche ? La réponse est non. En effet, comme discuté dans la section 3.4.7, le développement de la plateforme à destination des experts de la Ville a été interrompu. De ce fait, les visualisations proposées ne sont pas complètes. Des pistes d'amélioration sont décrites dans le

chapitre 4. Cette plateforme ne donne donc pas de réponse exhaustive à la première partie de la question de recherche. Par contre, la plateforme à destination des agents de la ville, bien qu'imparfaite, donne une proposition de réponse intéressante à la deuxième partie de la question de recherche. De plus, l'état de l'art, et plus particulièrement la table 2.1, synthétise la littérature par rapport à la visualisation du trafic et explique quelles techniques de visualisation sont pertinentes en fonction de l'objectif poursuivi.

Ce qui démarque la solution proposée des cas décrits dans l'état de l'art est l'utilisation de données hétérogènes. En effet, la plupart des exemples cités dans l'état de l'art n'utilisent qu'une seule source de données. La solution proposée a de son côté été articulée autour de plusieurs types de données provenant de différentes sources. La plateforme à destination des agents de la Ville est, à ce niveau, la plus représentative. En effet, elle regroupe des données provenant d'API Google, des données KML, les lieux de travail fournis par la Ville ainsi que les différentes constantes utilisées dans les calculs des gains et de la pollution.

Les contributions pratiques sont toutes les contributions qui apportent une plus-value à la ville de Charleroi. La première est l'inventaire des données disponibles qui est décrit dans la section 3.2. Cet inventaire permet à la Ville d'avoir une idée plus précise des données qu'elle a à sa disposition. La deuxième est l'état de l'art qui discute de plusieurs familles de techniques de visualisation de trafic et puis les compare avec les différents objectifs de visualisation de la mobilité dans la table 2.1. Cette table peut être utilisée par la Ville afin de définir quelles techniques utiliser en fonction de l'objectif poursuivi. Tout ceci donne des outils à la Ville qui peuvent l'aider à s'inscrire dans une stratégie de mobilité intelligente qu'elle souhaite atteindre. Enfin, la troisième contribution est la solution proposée en elle-même et plus particulièrement la plateforme à destination des agents de la Ville.

## Chapitre 4

# Perspectives d'amélioration

### 4.1 Introduction

Ce chapitre s'intéresse aux différentes modifications à apporter aux deux plateformes développées dans le chapitre 3. La section 4.2 se penche sur les visualisations à destination des experts de la Ville et discute des solutions à mettre en place pour que ces visualisations puissent être utilisées par les experts en mobilité de la Ville. De son côté, la visualisation à destination des agents de la Ville est déjà utilisable. La section 4.3 présente donc différentes idées d'amélioration du produit.

### 4.2 Visualisations à destination des expert la Ville

Cette section présente des améliorations à apporter à la plateforme à destination de la ville. Ces améliorations visent à répondre aux limites définies dans la section 3.4.7.

#### 4.2.1 Intégration des visualisations dans QGIS

Comme discuté dans la section 3.4.7, une plateforme web comme celle développée pour la Ville n'est pas intéressante pour les personnes travaillant sur QGIS. En effet, ces personnes travaillent principalement sur ce logiciel et l'ajout de la solution proposée les forcerait à travailler sur deux plateformes différentes. Dans ces conditions, il est probable que la solution proposée ne soit pas utilisée.

Une solution pourrait être d'intégrer les visualisations dans QGIS. Cela nécessiterait un changement radical de l'architecture de la solution car elle ne serait plus une plateforme web à déployer sur les serveurs de la Ville mais plutôt un *plugin* à développer et à intégrer dans QGIS. Ce logiciel permet en effet à des développeurs d'implémenter des extensions développées en Python ou R. Ces extensions ajoutent de nouvelles fonctionnalités au logiciel. Cependant, avant de se lancer dans une telle solution, il faut vérifier s'il est possible de développer des visualisations au sein de ces extensions. Si ce n'est pas le cas, alors une autre possibilité serait de garder la plateforme web telle qu'elle est développée pour le moment en ajoutant une possibilité pour l'utilisateur d'exporter les visualisations sous forme d'un *shapefile* qui pourrait par après être utilisé directement dans QGIS.

#### 4.2.2 Ajout de nouvelles données

Pour le moment, la solution proposée ne génère des visualisations que pour les agents de la cité administrative. Une amélioration possible pourrait être d'implémenter une interface d'administration permettant d'ajouter, modifier ou supprimer des données afin de générer ces mêmes visualisations mais pour d'autres jeux de données que la cité administrative.

La grande force de cette interface d'administration serait que l'administrateur pourrait ajouter très facilement de nouvelles données. Un bouton permettrait de choisir un fichier CSV sur son ordinateur et de l'importer sur le serveur de la Ville et donc sur la plateforme.

L'administrateur devrait donc juste choisir quel fichier CSV il veut importer et la plateforme s'occuperait du reste, c'est-à-dire du *preprocessing* des données et de la génération des visualisations. L'administrateur pourrait aussi simplement supprimer des jeux de données, les visualiser dans un tableau et même les modifier s'il le souhaite. La figure 4.1 donne un exemple de ce à quoi pourrait ressembler cette interface d'administration.

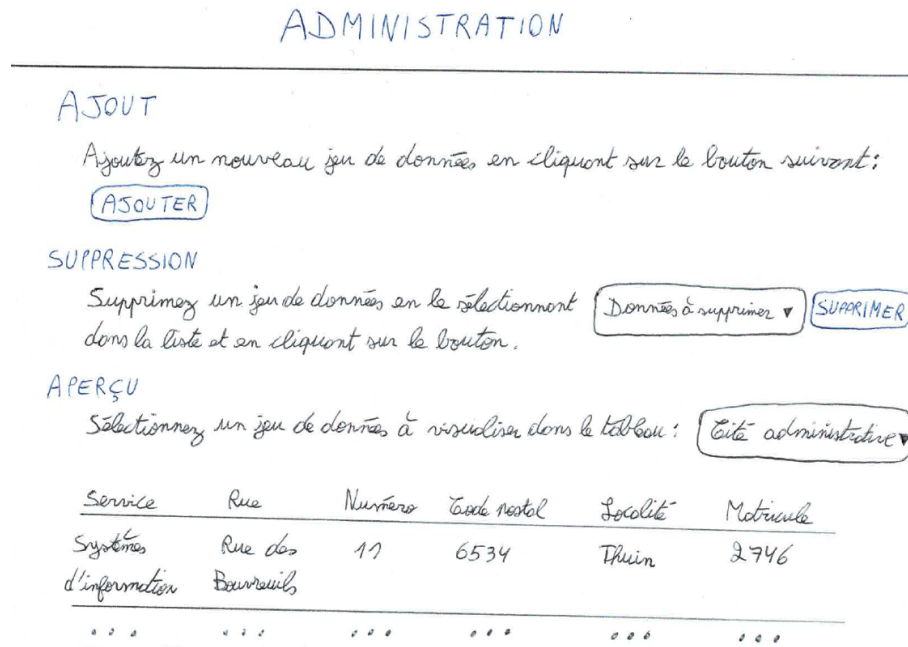


FIGURE 4.1 – Mockup présentant un exemple d'interface d'administration permettant à l'administrateur d'ajouter, supprimer et consulter les jeux de données utilisés pour générer les visualisations.

### 4.2.3 Affinage des visualisations

Le travail sur les visualisations a été rapidement interrompu. Ces visualisations ne sont pas complètes et il serait intéressant de continuer à les développer afin de les affiner le plus possible. Un premier exemple est l'ajout de *popup* présentant les données d'un *cluster* lorsque l'utilisateur clique sur celui-ci. Ces données pourraient par exemple être le nombre de domiciles qui caractérisent le flux. Un deuxième exemple est l'ajout de couches de données par dessus les différentes visualisations comme des couches KML. De manière générale, de l'interaction peut être ajoutée dans les deux visualisations afin d'aider au mieux l'utilisateur à savoir exactement ce qu'il analyse.

## 4.3 Visualisation à destination des agents de la Ville

Cette section propose différentes améliorations à apporter à la plateforme à destination des agents, que ce soit l'implémentation de nouvelles fonctionnalités ou encore une réflexion sur l'évolution générale de la plateforme.

### 4.3.1 Ajout d'une interface d'administration

La plateforme implémentée ne propose pour l'instant qu'une partie *frontend* qui reprend l'ensemble de la visualisation proposée aux agents. Cependant il serait intéressant d'y ajouter un *backend* qui implémenterait notamment un système d'authentification, ce qui permettrait par la suite de développer une interface d'administration accessible uniquement par un ou plusieurs administrateurs de la plateforme. L'ajout d'une telle interface d'administration répond à un retour important des clients discuté dans la section 3.4.7 : le manque d'aspects

paramétrables dans la solution.

Cette interface d'administration devrait être la plus simple possible afin que l'administrateur ne soit pas forcément un informaticien. Elle permettrait à la Ville de changer les différentes valeurs choisies lors du développement sans devoir passer par le code. Par exemple, il serait possible de modifier les valeurs de pollution et de coût au kilomètre pour un type de transport ou encore de modifier les distances pivots définissant les moyens de transport proposés. Cette interface d'administration permettrait également de modifier la liste des lieux de travail proposée à l'utilisateur ainsi que les caractéristiques de ceux-ci comme par exemple la présence ou non d'un abonnement de parking de 120€.

#### 4.3.2 Ajout d'un paramétrage de la solution par l'utilisateur

Pour le moment, l'utilisateur n'influence les résultats de la plateforme qu'avec l'insertion de l'adresse de son domicile et le choix de son lieu de travail. La plateforme est intéressante car elle donne des résultats différents en fonction du lieu d'habitation de l'agent. Cependant, comme l'a fait remarquer le client, ces résultats sont peut-être trop génériques et ne s'adaptent pas forcément aux desiderata de tous les agents qui vont utiliser la plateforme. En effet, si un agent veut absolument se déplacer en vélo alors qu'il habite à plus de 5 km de son lieu de travail, la plateforme ne lui propose pas de trajet en vélo.

Une solution à ce problème pourrait être de rendre les résultats plus personnalisables pour chacun des agents en complexifiant le formulaire proposé pour y intégrer leurs envies. Pour reprendre le même exemple, il pourrait notamment y avoir une *checkbox* permettant à l'utilisateur de demander au système de lui proposer un trajet en vélo.

Attention cependant à ne pas trop compliquer le formulaire. En effet, la principale force de la plateforme actuelle est sa simplicité d'utilisation : l'utilisateur ne doit pas réaliser beaucoup d'actions pour que le système retourne un résultat. L'ajout de nouvelles options dans le formulaire pourrait rendre la plateforme moins ergonomique et donc moins facile d'utilisation. Une solution pourrait être de proposer tout d'abord une version simplifiée du formulaire, c'est-à-dire le formulaire actuel demandant uniquement l'adresse de l'agent et son lieu de travail avec une possibilité d'afficher, si l'agent le souhaite, des options supplémentaires.

#### 4.3.3 Changement de la cible

Les différentes démonstrations du produit faites devant le client ont donné à celui-ci une compréhension profonde de l'ensemble des fonctionnalités de la plateforme. Cela lui a permis d'imaginer son évolution dans les mois et années à venir. La plateforme a été développée pour une cible particulière que sont les 3500 agents. L'idée de la Ville est de proposer une plateforme à destination de l'ensemble de la population de Charleroi. L'objectif resterait toujours le même, c'est-à-dire inciter les habitants à changer leurs habitudes de transport pour se rendre à leur lieu de travail.

Ce changement de cible implique une augmentation considérable du nombre d'utilisateurs potentiels qui passerait de 3500 à environ 400 000. Outre les changements à effectuer dans le code de l'application afin que celle-ci soit *scalable*, ce grand nombre d'utilisateur est une opportunité pour la Ville de recueillir des informations sur les habitudes de transport. La plateforme, que ce soit avec l'une ou l'autre cible, ne garde pas en mémoire les recherches effectuées par les utilisateurs. Une manière de recueillir des informations pourrait être de proposer aux utilisateurs de spécifier à la plateforme leur moyen de transport domicile-travail, et ce régulièrement. Cela permettrait à celle-ci de donner des statistiques d'utilisation des moyens de transports ainsi qu'une évolution à travers le temps. La Ville pourrait alors avoir des informations sur les habitudes de ses citoyens en matière de transport.

#### 4.3.4 Affinage de la visualisation

Comme pour les visualisations à destination de la Ville, la visualisation à destination des agents peut être affinée de plusieurs manières différentes. Tout d'abord en ajoutant la

possibilité à l'utilisateur de télécharger et imprimer un fichier PDF rassemblant les informations fournies par la visualisation. Dans certaines circonstances, une version papier est plus indiquée que la version web. Si par exemple l'agent décide de se rendre à son lieu de travail en vélo mais qu'il ne connaît pas bien le trajet proposé par la plateforme, il pourra imprimer la carte et ainsi avoir le trajet dans sa poche.

Enfin, l'horaire proposé en transport en commun est l'horaire qui permet à l'utilisateur d'arriver au maximum à 8h30 le matin et de partir du travail à partir de 17h. Cependant ce trajet n'est peut-être pas le plus rapide. Par exemple si la plateforme propose à un agent travaillant à l'hôtel de ville de prendre un train L à 7h45 qui arrive à la gare de Charleroi Sud à 8h20 et puis de marcher 10 minutes, l'agent arrive bien à 8h30 à l'hôtel de ville. Mais il existe peut-être un train IC partant à 7h55 et arrivant à 8h10 qui pourrait être privilégié par l'agent sans que ce trajet ne soit proposé par la plateforme. C'est pourquoi il serait intéressant de proposer un ensemble de trajets de transports en commun dans une tranche horaire définie afin que l'agent ait le choix entre plusieurs transports en communs.



## Chapitre 5

# Conclusion

La ville de Charleroi a un certain nombre de besoins de mobilité à satisfaire. L'un de ces besoins concerne les déplacements domicile-travail de l'ensemble des personnes travaillant pour la ville, appelés agents. Il peut être divisé en deux parties. Tout d'abord, la ville de Charleroi aimerait connaître les habitudes de déplacement domicile-travail de ses agents afin de répondre à un questionnaire fédéral sur les déplacements de ses travailleurs. Ensuite, elle aimerait pousser ses agents à utiliser des moyens de transport plus durables pour se rendre à leur lieu de travail.

Afin de répondre à ce double besoin, un stage s'est déroulé durant le deuxième trimestre de l'année académique 2019-2020 à l'hôtel de ville de Charleroi. Le stage a notamment permis une exploration d'un ensemble de techniques de visualisation afin de définir comment elles pourraient répondre à ce double besoin. Il a aussi permis de réaliser plusieurs contributions dont la ville de Charleroi peut profiter.

Une première contribution est le développement de deux plateformes répondant à la question de recherche et pour lesquelles une évaluation avec les utilisateurs a pu être réalisée. La première plateforme est destinée aux experts en mobilité de la ville de Charleroi. Son objectif est de fournir des informations sur les déplacements des agents de la ville. Cette plateforme est un prototype, elle ne propose donc que des visualisations qui se basent sur les agents qui travailleront à la cité administrative. Malheureusement, son développement a été arrêté durant le stage. En effet l'utilisation d'une plateforme web s'éloigne trop des habitudes des experts qui utilisent surtout des logiciels comme QGIS. Or il n'était pas possible dans le temps imparti d'implémenter des visualisations intégrées directement dans QGIS.

La deuxième plateforme est destinée à l'ensemble des agents de la ville de Charleroi et leur propose une visualisation qui répond à la deuxième partie du besoin. Elle informe les agents sur les différents moyens de transport possibles pour se rendre à leur lieu de travail et les pousse à utiliser des moyens de transport plus durables que la voiture. L'objectif sous-jacent est double : réduire les problèmes de congestion dûs aux nombreuses voitures présentes en ville et diminuer l'empreinte carbone des agents.

Une deuxième contribution concerne l'inventaire des données dont la ville de Charleroi dispose et de ses besoins. En effet, un lien a été créé entre les données dont la ville dispose et les problèmes de mobilité auxquels elle fait face. Les difficultés liées à chacun des besoins ont été discutées et montrent pourquoi, à l'heure actuelle, certains besoins ne peuvent être satisfaits avec les données disponibles. Ces informations sont très intéressantes pour la ville de Charleroi car elles peuvent l'aider à orienter ses futures démarches d'acquisition de données notamment.

La dernière contribution est l'état de l'art. Au-delà de la synthèse d'un ensemble de techniques de visualisation du trafic, cet état de l'art propose un croisement entre ces techniques, en fonction des données utilisées, et les objectifs de visualisation dans le domaine de la mobilité. Ce croisement est doublement intéressant. D'un point de vue purement théorique, des chercheurs pourraient le compléter avec d'autres techniques ou l'utiliser pour développer de

nouvelles classifications. D'un point de vue pratique, la ville de Charleroi pourrait se servir de ce tableau comme d'un outil qui la guiderait vers la réalisation de futurs travaux qui répondraient à des problématiques de mobilité. Tout ceci aiderait la Ville à s'inscrire dans une démarche de mobilité intelligente.

Les deux outils développés pour répondre à la question de recherche pourraient être améliorés de différentes manières. Pour le premier outil, l'amélioration la plus importante serait de changer la structure de la solution afin d'intégrer les visualisations dans QGIS. Cette intégration permettrait aux utilisateurs de ne pas devoir changer leurs habitudes de travail. Pour le deuxième outil, une interface d'administration pourrait être implémentée afin que la ville de Charleroi puisse paramétrer plus facilement les résultats de la solution. Pour le moment, ces paramètres doivent être modifiés dans le code. Une interface d'administration ajouterait une certaine flexibilité à la solution.

# Bibliographie

- ANDRIENKO, G., ANDRIENKO, N., CHEN, W., MACIEJEWSKI, R. et ZHAO, Y. (2017). Visual analytics of mobility and transportation : State of the art and further research directions. IEEE Transactions on Intelligent Transportation Systems, 18(8):2232–2249.
- ANDRIENKO, G., ANDRIENKO, N., FUCHS, G. et WOOD, J. (2016a). Revealing patterns and trends of mass mobility through spatial and temporal abstraction of origin-destination movement data. IEEE transactions on visualization and computer graphics, 23(9):2120–2136.
- ANDRIENKO, N. et ANDRIENKO, G. (2013). Visual analytics of movement : An overview of methods, tools and procedures. Information Visualization, 12(1):3–24.
- ANDRIENKO, N., ANDRIENKO, G. et RINZIVILLO, S. (2016b). Leveraging spatial abstraction in traffic analysis and forecasting with visual analytics. Information Systems, 57:172–194.
- AZZAM, T., EVERGREEN, S., GERMUTH, A. A. et KISTLER, S. J. (2013). Data visualization and evaluation. New Directions for Evaluation, 2013(139):7–32.
- BOSTOCK, M., OGIEVETSKY, V. et HEER, J. (2011). D<sup>3</sup> data-driven documents. IEEE transactions on visualization and computer graphics, 17(12):2301–2309.
- BOYANDIN, I., BERTINI, E. et LALANNE, D. (2010). Visualizing the world’s refugee data with jflowmap.
- BUCHIN, K., SPECKMANN, B. et VERBEEK, K. (2011). Flow map layout via spiral trees. IEEE transactions on visualization and computer graphics, 17(12):2536–2544.
- BYRON, L. et WATTENBERG, M. (2008). Stacked graphs–geometry & aesthetics. IEEE transactions on visualization and computer graphics, 14(6):1245–1252.
- CHEN, W., GUO, F. et WANG, F.-Y. (2015). A survey of traffic data visualization. IEEE Transactions on Intelligent Transportation Systems, 16(6):2970–2984.
- CRUZ, P. (2012). Data lenses. <http://pmcruz.com/information-visualization/data-lenses>. Consulté le 16 mars 2020.
- CUI, W., ZHOU, H., QU, H., WONG, P. C. et LI, X. (2008). Geometry-based edge clustering for graph visualization. IEEE Transactions on Visualization and Computer Graphics, 14(6):1277–1284.
- ERSOY, O., HURTER, C., PAULOVICH, F., CANTAREIRO, G. et TELEA, A. (2011). Skeleton-based edge bundling for graph visualization. IEEE transactions on visualization and computer graphics, 17(12):2364–2373.
- ESRI (1998). Esri shapefile technical description. <https://www.esri.com/library/whitepapers/pdfs/shapefile.pdf>, Consulté le 26 février 2020.
- FANG, T. B. et LU, Y. (2011). Constructing a near real-time space-time cube to depict urban ambient air pollution scenario. Transactions in GIS, 15(5):635–649.
- FERREIRA, N., POCO, J., VO, H. T., FREIRE, J. et SILVA, C. T. (2013). Visual exploration of big spatio-temporal urban data : A study of new york city taxi trips. IEEE transactions on visualization and computer graphics, 19(12):2149–2158.

- FORTUNATI, L. et TAIPALE, S. (2017). Mobilities and the network of personal technologies : Refining the understanding of mobility structure. Telematics and Informatics, 34(2):560–568.
- GONDIM, H., do NASCIMENTO, H. et REILLY, D. (2014). Visualizing large scale vehicle traffic network data a survey of the state-of-the-art. In 2014 International Conference on Information Visualization Theory and Applications (IVAPP), pages 337–346. IEEE.
- GUO, D. (2009). Flow mapping and multivariate visualization of large spatial interaction data. IEEE Transactions on Visualization and Computer Graphics, 15(6):1041–1048.
- GUO, D., CHEN, J., MAC EACHREN, A. M. et LIAO, K. (2006). A visualization system for space-time and multivariate patterns (vis-stamp). IEEE transactions on visualization and computer graphics, 12(6):1461–1474.
- GUO, D. et ZHU, X. (2014). Origin-destination flow data smoothing and mapping. IEEE Transactions on Visualization and Computer Graphics, 20(12):2043–2052.
- GUO, D., ZHU, X., JIN, H., GAO, P. et ANDRIS, C. (2012). Discovering spatial patterns in origin-destination mobility data. Transactions in GIS, 16(3):411–429.
- GUO, H., WANG, Z., YU, B., ZHAO, H. et YUAN, X. (2011). Tripvista : Triple perspective visual trajectory analytics and its application on microscopic traffic data at a road intersection. In 2011 IEEE Pacific Visualization Symposium, pages 163–170. IEEE.
- HARROWER, M. et BREWER, C. A. (2003). Colorbrewer. org : an online tool for selecting colour schemes for maps. The Cartographic Journal, 40(1):27–37.
- JONES, P. (2014). The evolution of urban mobility : The interplay of academic and policy perspectives. IATSS research, 38(1):7–13.
- KAPLER, T. et WRIGHT, W. (2005). Geotime information visualization. Information visualization, 4(2):136–146.
- KLOECKL, K., CHEN, X., SOMMER, C., RATTI, C. et BIDERMAN, A. (2016). Trains of data. <http://senseable.mit.edu/trainsofdata/>. Consulté le 16 mars 2020.
- KRAAK, M.-J. (2003). The space-time cube revisited from a geovisualization perspective. In Proc. 21st International Cartographic Conference, pages 1988–1996. Citeseer.
- KRIEGLER, H.-P., KRÖGER, P., SANDER, J. et ZIMEK, A. (2011). Density-based clustering. Wiley Interdisciplinary Reviews : Data Mining and Knowledge Discovery, 1(3):231–240.
- KUCHER, K. et KERREN, A. (2015). Text visualization techniques : Taxonomy, visual survey, and community insights. In 2015 IEEE Pacific Visualization Symposium (PacificVis), pages 117–121. IEEE.
- LAMPE, O. D. et HAUSER, H. (2011). Interactive visualization of streaming data with kernel density estimation. In 2011 IEEE Pacific visualization symposium, pages 171–178. IEEE.
- LAROUSSE (s.d.). Mobilité. <https://www.larousse.fr/dictionnaires/francais/mobilit%C3%A9/51890?q=mobilit%C3%A9#51768>. Consulté le 8 mars 2020.
- LIKAS, A., VLASSIS, N. et VERBEEK, J. J. (2003). The global k-means clustering algorithm. Pattern recognition, 36(2):451–461.
- LIU, S., PU, J., LUO, Q., QU, H., NI, L. M. et KRISHNAN, R. (2013). Vait : A visual analytics system for metropolitan transportation. IEEE Transactions on Intelligent Transportation Systems, 14(4):1586–1596.
- MALIENE, V., GRIGONIS, V., PALEVIČIUS, V. et GRIFFITHS, S. (2011). Geographic information system : Old principles with new capabilities. Urban Design International, 16(1):1–6.
- MCDANIEL, M. A. et EINSTEIN, G. O. (1986). Bizarre imagery as an effective memory aid : The importance of distinctiveness. Journal of experimental psychology : Learning, memory, and cognition, 12(1):54.

- MIT (1996). Brain processing of visual information. <https://news.mit.edu/1996/visualprocessing>.
- MUNZNER, T. (2014). Visualization analysis and design. CRC press.
- O’SULLIVAN, D. et UNWIN, D. (2014). Geographic information analysis. John Wiley & Sons.
- PACK, M. L., WONGSUPHASAWAT, K., VANDANIKER, M. et FILIPPOVA, D. (2009). Ice-visual analytics for transportation incident datasets. In 2009 IEEE International Conference on Information Reuse & Integration, pages 200–205. IEEE.
- PHAN, D., XIAO, L., YEH, R. et HANRAHAN, P. (2005). Flow map layout. In IEEE Symposium on Information Visualization, 2005. INFOVIS 2005., pages 219–224. IEEE.
- PU, J., LIU, S., DING, Y., QU, H. et NI, L. (2013). T-watcher : A new visual analytic system for effective traffic surveillance. In 2013 IEEE 14th International Conference on Mobile Data Management, volume 1, pages 127–136. IEEE.
- RIVADENEIRA, A. W., GRUEN, D. M., MULLER, M. J. et MILLEN, D. R. (2007). Getting our head in the clouds : toward evaluation studies of tagclouds. In Proceedings of the SIGCHI conference on Human factors in computing systems, pages 995–998.
- ROBINSON, A. H. et RANDALL, D. (1969). Sale. 1969. elements of cartography.
- SHNEIDERMAN, B. (1996). The eyes have it : A task by data type taxonomy for information visualizations. In Proceedings 1996 IEEE symposium on visual languages, pages 336–343. IEEE.
- ST. JOHN, M., COWEN, M. B., SMALLMAN, H. S. et OONK, H. M. (2001). The use of 2d and 3d displays for shape-understanding versus relative-position tasks. Human Factors, 43(1):79–98.
- STEPHEN, D. M. et JENNY, B. (2017). Automated layout of origin–destination flow maps : Us county-to-county migration 2009–2013. Journal of Maps, 13(1):46–55.
- STEVENS, S. S. et al. (1946). On the theory of scales of measurement.
- TOMINSKI, C., SCHUMANN, H., ANDRIENKO, G. et ANDRIENKO, N. (2012). Stacking-based visualization of trajectory attribute data. IEEE Transactions on visualization and Computer Graphics, 18(12):2565–2574.
- TORGUSON, J. S. (2016). Choropleth map. International Encyclopedia of Geography : People, the Earth, Environment and Technology : People, the Earth, Environment and Technology, pages 1–9.
- UNITED NATIONS (2014). World’s population increasingly urban with more than half living in urban areas,. Rapport technique, (New-York, USA : United Nations, 2014). Accessed February 27, 2020.
- UNWIN, D. J. (1981). Introductory spatial analysis, volume 748. Taylor & Francis.
- URRY, J. (2007). Mobilities polity press.
- Van den ELZEN, S. et VAN WIJK, J. J. (2014). Multivariate network exploration and presentation : From detail to overview via selections and aggregations. IEEE Transactions on Visualization and Computer Graphics, 20(12):2310–2319.
- WARE, C. (2008). Visual Thinking for Design. Morgan Kaufmann.
- WARE, C. (2020). Information visualization : perception for design. Morgan Kaufmann.
- WENZHAN, W., YEXUN, Z., YUEYING, P. et HONGBO, Z. (2003). Mathematical definitions of point, line and area symbols in cartography. Geo-spatial Information Science, 6(3):62–65.
- WONG, P. C. et THOMAS, J. (2004). Visual analytics. IEEE Computer Graphics and Applications, (5):20–21.

- WOOD, J., DYKES, J. et SLINGSBY, A. (2010). Visualisation of origins, destinations and flows with od maps. The Cartographic Journal, 47(2):117–129.
- WOOD, J., SLINGSBY, A. et DYKES, J. (2011). Visualizing the dynamics of london’s bicycle-hire scheme. Cartographica : The International Journal for Geographic Information and Geovisualization, 46(4):239–251.
- ZHOU, H., XU, P., YUAN, X. et QU, H. (2013). Edge bundling in information visualization. Tsinghua Science and Technology, 18(2):145–156.